金融经济学

课程说明: 本课程是金融学等专业的硕士必修课程。它讲授理论金融经济学基础,并对以前出现在《投资学》、《公司财务》等课程中来源于现代理论金融经济学中的各种原理、方法纳入统一的理论框架中,使学生对它们有更深刻的理解。现代金融经济学中最重要的是无套利假设和一般经济均衡框架。所研究的中心问题是在不确定的金融市场环境下对金融资产定价。中心结果是资产定价基本定理。投资组合分析、资本资产定价模型 (CAPM)、套利定价理论 (APT)、期权定价理论、市场有效性理论、利率期限结构等都是这一框架中围绕这一中心结果的组成部分。作为硕士课程,本课程力图在尽可能少的数学要求下阐明基本经济思想。对数学要求较高的金融经济学理论将在高级金融经济学专题中讲述。

主讲教师: 史树中　　电话: 62757074　　e-mail: `shisz@gsm.pku.edu.cn`

教材: 自编

教学大纲:

1. 金融市场的不确定性。收益与风险。(以随机变量来刻划不确定的未来收益,风险用随机变量的方差来刻划。提出理论的基本数学模型。相当多的内容可以通过有限状态的古典概率论来描述。)

2. 金融经济学的理论与方法。(金融市场价格变化现象 − 统计分析 − 理论框架之间的关系。)

3. 一般经济均衡与无套利假设。资产定价基本定理。(新古典主义理论框架。基本公理体系。)

4. Modigliani-Miller 定理 (MMT)。(局限于二期模型。)

5. Markowitz 投资组合理论。资本资产定价模型 (CAPM)。(需要初等概率论,线性代数,和初等微积分。)

6. 套利定价理论 (APT)。(由渐近无套利假设来推导。阐明与 APT 的初等叙述之间的关系。)

7. 不确定状况下的决策。von Neumann-Morgenstein 效用函数。(用期望效用函数刻划不确定状况下的决策的依据。不确定状况下的一般经济均衡框架。需要初等概率论。)

8. CAPM 和 APT 的一般均衡理论模型。((投资(证券组合选择) − 消费模型。CAPM 和 APT 的一般经济均衡根据。完全市场与不完全市场的一般均衡理论介绍。)

9. 市场有效性理论。(市场有效性与无套利假设。市场有效性的数学表达与检验。)
10. Black-Scholes 期权定价理论。(Cox-Ross-Robinstein 二叉树方法。随机分析初步，Black-Scholes 公式的推导。Black-Scholes 方程。了解随机分析和偏微分方程的初步概念，但不作深究。)
11. 债券的利率期限结构。(利率期限结构的概念。Black-Scholes 方程对利率期限结构的应用。各种特殊模型。)
12. 动态 CAPM 和 APT。(介绍多时期模型和连续模型。)

数学预备知识: 多元微积分，线性代数，初等概率论与数理统计。

相关课程: 投资学，公司财务，金融工程，金融计量经济学，高等金融经济学专题。

考查方式: 作业 (30%)；完成论文一篇 (30%)；期末考试 (40%)。

参考书:

1. Jarrow, R. A. V. Maksimovic & W. T. Ziemba (eds), *Finance*, Elsevier, Amsterdam, 1995, Chaps., 1, 2, 4, and 9.
2. Jarrow, R. A., Finance Theory, Prentice Hall, Englewood Cliffs, NJ, 1988.
3. Huang, Chi-fu & R. H. Litzenberger, Foundations for Financial Economics, Prentice Hall, Englewood Cliffs, NJ, 1988.
4. Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.
5. Cochrane, J. H., 2001, Asset Pricing, Princeton University Press, Princeton.

金融经济学
十讲

史树中

北京大学金融数学与金融工程研究中心
北京大学光华管理学院 金融系

2003 年 2－7 月

从数理经济学到数理金融学的百年回顾[1]

(附“补记”(2003 年 2 月))

(代引言)

一般经济均衡理论和数学公理化

1874 年 1 月，在瑞士洛桑大学拥有教席的法国经济学家瓦尔拉斯 (L. Warlas, 1834～1910) 发表了他的论文《交换的数学理论原理》，首次公开他的一般经济均衡理论的主要观点。虽然数理经济学的创始人通常认为是法国数学家、经济学家和哲学家古诺 (A. A. Cournot, 1801～1877)，他在 1838 年出版了《财富理论的数学原理研究》一书，但是对今日的数理经济学影响最大的是瓦尔拉斯的一般经济均衡理论。尤其是，直到现在为止，一般经济均衡理论仍然是唯一的对经济整体提出的理论。

所谓一般经济均衡理论大致可以这样来简述：在一个经济体中有许多经济活动者，其中一部分是消费者，一部分是生产者。消费者追求消费的最大效用，生产者追求生产的最大利润，他们的经济活动分别形成市场上对商品的需求和供给。市场的价格体系会对需求和供给进行调节，最终使市场达到一个理想的一般均衡价格体系。在这个体系下，需求与供给达到均衡，而每个消费者和每个生产者也都达到了他们的最大化要求。瓦尔拉斯把这一思想表达为这样的数学问题：假定市场上一共有 l 种商品，每一种商品的供给和需求都是这 l 种商品的价格的函数。于是这 l 种商品的供需均衡就得到 l 个方程。但是价格需要有一个计量单位，或者说实际上只有各种商品之间的比价才有意义，因而，这 l 种商品的价格之间只有 $l-1$ 种商品的价格是独立的。为此，瓦尔拉斯又加入了一个财务均衡的关系，即所有商品供给的总价值应该等于所有商品需求的总价值。这一关系目前就称为“瓦尔拉斯法则”，它被用来消去一个方程。这样，最后瓦尔拉斯就认为，他得到了求 $l-1$ 种商品价格的 $l-1$ 个方程所组成的方程组。按照当时已为人们熟知的线性方程组理论，这个方程组有解，其解就是一般均衡价格体系。

瓦尔拉斯当过工程师，也专门向人求教过数学。这使他能把他的一般经济均衡的思想表达成数学形式。但是他的数学修养十分有限。

---

[1] 原载《科学》**52: 6** (2000), 29–33.

事实上，他提出的上述“数学论证”在数学上是站不住脚的。这是因为如果方程组不是线性的，那么方程组中的方程个数与方程是否有解就没有什么直接关系。这样，从数学的角度来看，长期来，瓦尔拉斯的一般经济均衡体系始终没有坚实的基础。这个问题经过数学家和经济学家们 80 年的努力，才得以解决。其中包括大数学家冯・诺依曼 (J. von Neumann, 1903～1957)，他曾在 20 世纪 30 年代投身到一般经济均衡的研究中去，并因此提出他的著名的经济增长模型；还包括 1973 年诺贝尔经济学奖获得者列昂节夫 (W. Leontiev, 1906～1999)，他在 20 世纪 30 年代末开始他的投入产出方法的研究，这种方法在实质上是一个一般经济均衡的线性模型。分别获得 1970 年和 1972 年诺贝尔经济奖的萨缪尔森 (P. Samuleson, 1915～) 和希克斯 (J. R. Hicks, 1904～1989)，也是因他们用数学方式研究一般经济均衡体系而著称。而最终在 1954 年给出一般经济均衡存在性的严格证明的是阿罗 (K. Arrow, 1921～) 和德布鲁 (G. Debreu, 1921～)。他们对一般经济均衡问题给出了富有经济含义的数学模型，利用 1941 年日本数学家角谷静夫 (Kakutani Shizuo, 1911～) 对 1911 年发表的荷兰数学家布劳维 (L. E. J. Brouwer, 1881～1967) 提出的不动点定理的推广，才给出一般经济均衡价格体系的存在性证明。他们两人也因此先后于 1972 年和 1983 年获得诺贝尔经济学奖。

阿罗和德布鲁都以学习数学开始他们的学术生涯。阿罗有数学的学士和硕士学位，德布鲁则完全是主张公理化方法的法国布尔巴基 (Bourbaki) 学派培养出来的数学家。他们两人是继冯・诺依曼以后，最早在经济学中引入数学公理化方法的学者。阿罗在 1951 年出版的《社会选择与个人价值》一书中，严格证明了满足一些必要假设的社会决策原则不可能不恒同于“某个人说了算”的“独裁原则”。这就是著名的阿罗不可能性定理。而德布鲁则是在他与阿罗一起证明的一般经济均衡存在定理的基础上，把整个一般经济均衡理论严格数学公理化，形成他于 1959 年出版的《价值理论》一书。这本 114 页的小书，今天已被认为是现代数理经济学的里程碑。

经济学为什么需要数学公理化方法是一个始终存在争论的问题。对于这个问题，德布鲁的回答是：“坚持数学严格性，使公理化已经不止一次地引导经济学家对新研究的问题有更深刻的理解，并使适合这些问题的数学技巧用得更好。这就为向新方向开拓，建立了一个可靠的基地。它使研究者从必须推敲前人工作的每一细节的桎梏中脱身出来。严格性无疑满足了许多当代经济学家的智力需要，因此，他们为了自身的原因而追求它，但是作为有效的思想工具，它也是理论的标志。” ([4]) 在这样的意义下，我们才能正确理解现代数理经济学、数理金融学的发展究竟意味着什么。当然，这并非意味着通过对各种现象、实例、故事的描述、罗列、区分，使人们从中悟出许多哲理来的“文学文化”的认识方法不能认识经济学、金融学的一些方面。但是认为经济

学、金融学不需要用公理化方法架构的科学理论，而只需要对经济现实、金融市场察颜观色的经验，那将更不能认识经济学、金融学的本质。

从“华尔街革命”追溯到 1900 年

狭义的金融学是指金融市场的经济学。现代意义下的金融市场至少已有三百年以上的历史，它从一开始就是经济学的研究对象。但是现代金融学通常认为只有不到50年的历史。这50年也就是使金融学成为可用数学公理化方法架构的历史。从瓦尔拉斯－阿罗－德布鲁的一般经济均衡体系的观点来看，现代金融学的第一篇文献是阿罗于 1953 年发表的论文《证券在风险承担的最优配置中的作用》。在这篇论文中，阿罗把证券理解为在不确定的不同状态下有不同价值的商品。这一思想后来又被德布鲁所发展，他把原来的一般经济均衡模型通过拓广商品空间的维数来处理金融市场，其中证券无非是不同时间、不同情况下有不同价值的商品。但是后来大家发现，把金融市场用这种方式混同于普通商品市场是不合适的。原因在于它掩盖了金融市场的不确定性本质。尤其是其中隐含着对每一种可能发生的状态都有相应的证券相对应，如同每一种可能有的金融风险都有保险那样，与现实相差太远。

这样，经济学家们又为金融学寻求其他的数学架构。新的数学架构的现代金融学被认为是两次“华尔街革命”的产物。第一次“华尔街革命”是指 1952 年马科维茨 (H. Markowitz, 1927～) 的证券组合选择理论的问世。第二次“华尔街革命”是指 1973 年布莱克 (F. Black, 1938～1995)－肖尔斯 (M. Scholes, 1941～) 期权定价公式的问世。这两次“革命”的特点之一都是避开了一般经济均衡的理论框架，以致在很长时期内都被传统的经济学家们认为是“异端邪说”。但是它们又确实使以华尔街为代表的金融市场引起了“革命”，从而最终也使金融学发生根本改观。马科维茨因此荣获 1990 年的诺贝尔经济学奖，肖尔斯则和对期权定价理论作出系统研究的默顿 (R. Merton, 1944～) 一起荣获 1997 年的诺贝尔经济学奖。不幸的是布莱克于 1995 年早逝，没有与他们一起领奖。

马科维茨研究的是这样的一个问题：一个投资者同时在许多种证券上投资，那么应该如何选择各种证券的投资比例，使得投资收益最大，风险最小。对此，马科维茨在观念上的最大贡献在于他把收益与风险这两个原本有点含糊的概念明确为具体的数学概念。由于证券投资上的收益是不确定的，马科维茨首先把证券的收益率看作一个随机变量，而收益定义为这个随机变量的均值 (数学期望)，风险则定义为这个随机变量的标准差 (这与人们通常把风险看作可能有的损失的思想相

差甚远)。于是如果把各证券的投资比例看作变量，问题就归结为怎样使证券组合的收益最大、风险最小的数学规划。对每一固定收益都求出其最小风险，那么在风险－收益平面上，就可画出一条曲线，它称为组合前沿。马科维茨理论的基本结论就是：在证券允许卖空的条件下，组合前沿是一条双曲线的一支；在证券不允许卖空的条件下，组合前沿是若干段双曲线段的拼接。组合前沿的上半部称为有效前沿。对于有效前沿上的证券组合来说，不存在收益和风险两方面都优于它的证券组合。这对于投资者的决策来说自然有很重要的参考价值。

马科维茨理论是一种纯技术性的证券组合选择理论。这一理论当年是作为他在芝加哥大学的博士论文提出的。但在论文答辩时，它被另一位当时已享有盛名、后来也以他的货币主义而获得 1976 年诺贝尔经济学奖的弗里德曼 (M. Friedman, 1912～) 斥之为：“这不是经济学！”为此，马科维茨后来不得不引入以收益和风险为自变量的效用函数，来使他的理论纳入通常的一般经济均衡框架。马科维茨的学生夏普 (W. Sharpe, 1934～) 和另一些经济学家，则进一步在一般经济均衡的框架下，假定所有投资者都以这种效用函数来决策，而导出全市场的证券组合的收益率是有效的以及所谓资本资产定价模型 (Capital Asset Pricing Model, CAPM)。夏普因此与马科维茨一起荣获 1990 年的诺贝尔经济学奖。另一位 1981 年诺贝尔经济奖获得者托平 (J. Tobin, 1918～) 在对于允许卖空的证券组合选择问题的研究中，导出每一种有效证券组合都是一种无风险资产与一种特殊的风险资产的组合 (它称为二基金分离定理)，从而得出一些宏观经济方面的结论。

在 1990 年与马科维茨和夏普一起分享诺贝尔奖的另一位经济学家是新近刚去世的米勒 (M. Miller, 1923～2000)。他与另一位在 1985 年获得诺贝尔奖的莫迪利阿尼 (F. Modigliani, 1918～) 一起在 1958 年以后发表了一系列论文，探讨“公司的财务政策 (分红、债权/股权比等) 是否会影响公司的价值”这一主题。他们的结论是：在理想的市场条件下，公司的价值与财务政策无关。后来他们的这些结论就被称为莫迪利阿尼－米勒定理。他们的研究不但为公司理财这门新学科奠定了基础，并且首次在文献中明确提出无套利假设。所谓无套利假设是指在一个完善的金融市场中，不存在套利机会 (即确定的低买高卖之类的机会)。因此，如果两个公司将来的 (不确定的) 价值是一样的，那么它们今天的价值也应该一样，而与它们财务政策无关；否则人们就可通过买卖两个公司的股票来获得套利。达到一般经济均衡的金融市场显然一定满足无套利假设。这样，莫迪利阿尼－米勒定理与一般经济均衡框架是相容的。但是直接从无套利假设出发来对金融产品定价，则使论证大大简化。这就给人以启发，我们不必一定要背上沉重的一般经济均衡的十字架，从无套利假设出发就已经可为金融产品的定价得到许多结果。从此，金融经济学就开始以无套利假设作为出发点。

以无套利假设作为出发点的一大成就也就是布莱克－肖尔斯期权定价理论。所谓 (股票买入) 期权是指以某固定的执行价格在一定的期限内买入某种股票的权利。期权在它被执行时的价格很清楚，即：如果股票的市价高于期权规定的执行价格，那么期权的价格就是市价与执行价格之差；如果股票的市价低于期权规定的执行价格，那么期权是无用的，其价格为零。现在要问期权在其被执行前应该怎样用股票价格来定价？为解决这一问题，布莱克和肖尔斯先把模型连续动态化。他们假定模型中有两种证券，一种是债券，它是无风险证券，也是证券价值的计量基准，其收益率是常数；另一种是股票，它是风险证券，沿用马科维茨的传统，它也可用证券收益率的期望和方差来刻划，但是动态化以后，其价格的变化满足一个随机微分方程，其含义是随时间变化的随机收益率，其期望值和方差都与时间间隔成正比。这种随机微分方程称为几何布朗 (Brown) 运动。然后，利用每一时刻都可通过股票和期权的适当组合对冲风险，使得该组合变成无风险证券，从而就可得到期权价格与股票价格之间的一个偏微分方程，其中的参数是时间、期权的执行价格、债券的利率和股票价格的“波动率”。出人意料的是这一方程居然还有显式解。于是布莱克－肖尔斯期权定价公式就这样问世了。

但是与马科维茨的遭遇类似，布莱克－肖尔斯公式的发表也困难重重地经过好几年。与市场中投资人行为无关的金融资产的定价公式，对于习惯于用一般经济均衡框架对商品定价的经济学家来说很难接受。这样，布莱克和肖尔斯不得不直接到市场中去验证他们的公式。结果令人非常满意。有关期权定价实证研究结果先在 1972 年发表。然后再是理论分析于 1973 年正式发表。与此几乎同时的是芝加哥期权交易所也在 1973 年正式推出 16 种股票期权的挂牌交易 (在此之前期权只有场外交易)，使得衍生证券市场从此蓬蓬勃勃地发展起来。布莱克－肖尔斯公式也因此有数不清的机会得到充分验证，而使它成为人类有史以来应用最频繁的一个数学公式。

布莱克－肖尔斯公式的成功与默顿的研究是分不开的，后者甚至在把他们的理论深化和系统化上作出更大的贡献。默顿的研究后来被总结在 1990 年出版的《连续时间金融学》一书中。对金融问题建立连续时间模型也在近 30 年中成为金融学的中心。这如同连续变量的微分学在瓦尔拉斯时代进入经济学那样，尽管现实的经济变量极少是连续的，微分学能强有力地处理经济学中的最大效用问题；而连续变量的金融模型同样使强有力的随机分析更深刻地揭示金融问题的随机性。不过用连续时间模型来处理金融问题并非从布莱克－肖尔斯－默顿理论开始。 20 世纪 50 年代，萨缪尔森就已发现，一位几乎被人遗忘的法国数学家巴施里叶 (L. Bachelier, 1870～1946) 早在 1900 年已经在他的博士论文《投机理论》中用布朗运动来刻划股票的价格变化，并且这是

历史上第一次给出的布朗运动的数学定义，比人们熟知的爱因斯坦 (A. Einstein, 1880～1955) 1905 年的有关布朗运动的研究还要早。尤其是巴施里叶实质上已经开始研究期权定价理论，而布莱克－肖尔斯－默顿的工作其实都是在萨缪尔森的影响下，延续了巴施里叶的工作。这样一来，数理金融学的“祖师爷”就成了巴施里叶。对此，法国人很自豪，最近他们专门成立了国际性的“巴施里叶协会”。2000 年 6 月，协会在巴黎召开第一届盛大的国际“巴施里叶会议”，以纪念巴施里叶的论文问世 100 周年。

谁将是下一位金融的诺贝尔经济奖得主？

布莱克－肖尔斯公式的成功也是用无套利假设来为金融资产定价的成功。这一成功促使 1976 年罗斯 (S. A. Ross, 1944～) 的套利定价理论 (APT, Arbitrage Pricing Theory) 的出现。APT 是作为 CAPM 的替代物而问世的。CAPM 的验证涉及对市场组合是否有效的验证，但是这在实证上是不可行的。于是针对 CAPM 的单因素模型，罗斯提出目前被统称为 APT 的多因素模型来取代它。对此，罗斯构造了一个一般均衡模型，证明了各投资者持有的证券价值在市场组合中的份额越来越小时，每种证券的收益都可用若干基本经济因素来一致近似地线性表示。后来有人发现，如果仅仅需要对各种金融资产定价的多因素模型作出解释，并不需要一般均衡框架，而只需要线性模型假设和“近似无套利假设”：如果证券组合的风险越来越小，那么它的收益率就会越来越接近无风险收益率。这样，罗斯的 APT 就变得更加名符其实。从理论上来说，罗斯在其 APT 的经典论文中更重要的贡献是提出了套利定价的一般原理，其结果后来被称为“资产定价基本定理”。这条定理可表述为：无套利假设等价于存在对未来不确定状态的某种等价概率测度，使得每一种金融资产对该等价概率测度的期望收益率都等于无风险证券的收益率。1979 年罗斯还与考克斯 (J. C. Cox) 和鲁宾斯坦 (M. Rubinstein) 一起，利用这样的资产定价基本定理对布莱克－肖尔斯公式给出了一种简化证明，其中股票价格被设想为在未来若干时间间隔中越来越不确定地分叉变化，而每两个时间间隔之间都有上述的“未来收益的期望值等于无风险收益率”成立。由此得到期权定价的离散模型。而布莱克－肖尔斯公式无非是这一离散模型当时间间隔趋向于零时的极限。

这样一来，金融经济学就在很大程度上离开了一般经济均衡框架，而只需要从等价于无套利假设的资产定价基本定理出发。由此可以得到许多为金融资产定价的具体模型和公式，并且形成商学院学生学习“投资学”的主要内容。 1998 年米勒在德国所作的题为《金融学的历史》的报告中把这样的现象描述成：金融学研究被分流为经济系探讨

的“宏观规范金融学”和商学院探讨的“微观规范金融学”。这里的主要区别之一就在于是否要纳入一般经济均衡框架。同时，米勒还指出，在金融学研究中，“规范研究”与“实证研究”之间的界线倒并不很清晰。无论是经济系的“宏观规范”研究还是商学院的“微观规范”研究一般都少不了运用模型和数据的实证研究。不过由于金融学研究与实际金融市场的紧密联系，“微观规范”研究显然比“宏观规范”研究要兴旺得多。

至此，从数理经济学到数理金融学的百年回顾已可基本告一段落。正如米勒在上述报告中所说，回顾金融学的历史有一方便之处，就是看看有谁因金融学研究而获得诺贝尔经济奖。我们的回顾同样利用了这一点。恰好在本文发稿期间，传来 2000 年的诺贝尔经济奖的消息：获奖者为两位美国经济学家詹姆斯・赫克曼 (James J. Heckman, 1944～) 和丹尼尔・麦克法登 (Daniel L. McFadden, 1937～)，以表彰他们在与本文主题密切相关的微观计量经济学领域所作出的贡献。那么还有谁会因为其金融学研究在 21 世纪获得诺贝尔奖呢？

从我们的叙述中看来，似乎罗斯有较大希望。但是在米勒的报告中，他更加推崇他的芝加哥大学的同事法玛 (E. F. Fama)。法玛的成就首先是因为他在 20 世纪 60 年代末开始的市场有效性方面的研究。所谓市场有效性问题是指市场价格是否充分反映市场信息的问题。当金融商品定价已经建立在无套利假设的基础上时，对市场是否有效的实证检验就和金融理论是否与市场现实相符几乎成了一回事。这大致可以这样来说，如果金融市场的价格变化能通过布朗运动之类的市场有效性假设的检验，那么市场就会满足无套利假设。这时，理论比较符合实际，而对投资者来说，因为没有套利机会，就只能采取保守的投资策略。而如果市场有效性假设检验通不过，那么它将反映市场有套利机会，市场价格在一定程度上有可预测性，投资者就应该采取积极的投资策略。业间流行的股市技术分析之类就会起较大作用。这样，市场有效性的研究对金融经济学和金融实践来说就变得至关紧要。法玛在市场有效性的理论表述和实证研究上都有重大贡献。法玛的另一方面影响极大的重要研究是最近几年来，他与弗兰齐 (K. French) 等人对 CAPM 的批评。他们认为，以市场收益率来刻划股票收益率，不足以解释股票收益率的各种变化。他们建议，引入公司规模以及股票市值与股票帐面值的比作为新的解释变量。他们的一系列论文引起金融界非常热烈的争论，并且已开始被人们广泛接受。虽然他们的研究基本上还停留在计量经济学的层次，但势必会对数理金融学的结构产生根本的影响。

法玛的研究是金融学中典型的“微观规范”与实证的研究。至于“宏观规范”的研究，我们应该提到关于不完全市场的一般经济均衡理

论研究。由无套利假设得出的资产定价基本定理以及原有的布莱克－肖尔斯理论实际上只能对完全市场中的金融资产唯一定价。这里的完全市场是指作为定价出发点的基本资产 (无风险证券、标的资产等) 能使每一种风险资产都可以表达为它们的组合。实际情况自然不会是这样。关于不完全证券市场的一般经济均衡模型是拉德纳 (R. Radner) 于 1972 年首先建立的，他同时在对卖空有限制的条件下，证明了均衡的存在性。但是过了三年，哈特 (O. Hart) 举出一个反例，说明在一般情况下，不完全证券市场的均衡不一定存在。这一问题曾使经济学家们困惑很久。一直到 1985 年，达菲 (D. Duffie) 和夏弗尔 (W. Schafer) 指出，对于“极大多数”的不完全市场，均衡还是存在的。遗憾的是，他们同时还证明了，不完全市场的“极大多数”均衡都不能达到“资源最优配置”。这样的研究结果的经济学含义值得人们深思。达菲和夏弗尔的数学证明还使数学家十分兴奋，因为他们用到例如格拉斯曼 (Grassmann) 流形上的不动点定理那样的对数学家来说也是崭新的研究。此后的十几年，沿着这一思想发展出一系列与完全市场相对应的各种各样的反映金融市场的不完全市场一般均衡理论。在这方面也有众多贡献的麦基尔 (M. Magill) 和奎恩兹 (M. Quinzii) 已经在世纪末为这一主题写出厚厚的两卷专著。这些数理经济学家作为个人对诺贝尔经济奖的竞争力可能不如罗斯和法玛，但是不完全市场一般经济均衡作为数理经济学和数理金融学的又一高峰，则显然是诺贝尔经济学奖的候选者。

21 世纪的到来伴随着计算机和互联网络飞速发展。在这些高新技术的推动下，金融市场将进一步全球化、网络化。网上交易、网上支付、网上金融机构、网上清算系统等更使金融市场日新月异。毫无疑问，21 世纪的数理金融学将更以我们意想不到的面貌向我们走来。

参考文献

[**1**] 史树中，数学与经济，湖南教育出版社，1990。

[**2**] 瓦尔拉斯，纯粹经济学要义，蔡树柏译，商务印书馆，1989。

[**3**] 德布鲁，价值理论，刘勇、梁日杰译，北京经济学院出版社，1989。

[**4**] 德布鲁，数学思辩形式的经济理论，史树中译，《数学进展》，**17: 3** (1988), 251–259.

[**5**] Miller, M. H., The history of finance, An eyewitness account, Journal of Portfolio Management, Summer 1999, 95–101.

补记 (2003 年 2 月)

本文发表至今已经有两年。在这两年中诺贝尔经济学奖又颁发了两次，并且都与金融学有关。但是既没有颁给罗斯，也没有颁给法玛，更没有颁给不完全市场理论。看来人们认为经典的金融学已经告一段落，而非经典的金融学必须考虑比均衡、无套利等更有活力的因素。这类因素之一是金融市场中的信息传递，之二是人们在金融市场中的决策心理。它们正是 2001 年和 2002 年诺贝尔经济学奖的两个主题。

2001 年的诺贝尔经济学奖授予三位美国经济学家阿克洛夫 (George A. Akerlof, 1940～)，斯彭思 (A. Michael Spence, 1943～) 和斯蒂格利茨 (Joseph E. Stiglitz, 1943～) 以奖励他们对具有不对称信息的市场的分析。所谓具有不对称信息的市场是司空见惯的。当前中国令人深恶痛绝的假冒伪劣商品市场就是有不对称信息的市场：卖方做了手脚，买方则蒙在鼓里。在信息不对称的市场中，价格会发生畸变。或者劣等品卖了好价钱，也有可能优等品被贱卖。而金融市场中的信息尽管比其他商品市场更要透明，但仍然存在严重的信息不对称。尤其是像在我国那样的历史较短的证券市场中，大户操纵，散户跟风现象曾经比比届是。即使是像美国那样的成熟市场，近年来也出现了大做假账的“安然事件”之类。在这种情况下，用均衡和无套利来为金融资产定价显然是不合理的。于是市场有效性就成了大问题。

斯蒂格利茨的一项得奖工作就是针对市场有效性的。1980 年他与格罗斯曼 (Stanford J. Grossman) 一起提出有关市场有效性的“Grossman-Stiglit 悖论”：如果市场价格已经反映了所有有关的市场信息，那么经济活动者就没有必要去搜集市场信息；但是如果所有经济活动者都不去搜集市场信息，那么市场价格怎么可能反映所有有关的市场信息？这样，经典的市场有效性理论就受到了严重挑战。关于这一悖论的研究对金融经济学的影响极大。其主要解决方案是在市场的一般经济均衡模型中需要引入有成本的信息，引进掌握不同信息的交易者。这一来就走出了经典金融学的无套利框架。

2002 年的诺贝尔经济学奖授于美国－以色列心理学家卡尼曼 (Daniel Kahneman, 1934～) 和美国经济学家史密斯 (Vernon L. Smith, 1927～)，以奖励他们在实验经济学和行为经济学方面的开创性工作。对卡尼曼是奖励他“对把心理研究融入经济科学，特别是有关在不确定环境下人们的判断和决策，有完整见解。”对史密斯是奖励他“在经验经济分析中，特别是在备选市场机理研究中，建立了实验室试验。”这里与金融学关系比较密切的是卡尼曼的工作。

卡尼曼完全是位心理学家，但是他现在已经与另一位已故的心理学家特浮尔斯基 (Amos Tversky, 1937～1996) 被公认为是行为经济学的

倡导人。他们两人于 1979 年发表的论文已成为《计量经济学 (Econometrica)》有史以来被引证最多的经典。他们研究的问题是人们在不确定环境下的判断和决策。在此以前，人们运用的传统理论是冯·诺依曼和摩根斯特恩 (O. Morgenstern) 1944 年提出的期望效用函数理论。这一理论用数学公理化的方法证明，每个人在不确定环境下的决策可通过求他的一个效用函数的平均值（数学期望）的最大值来描述。虽然它在数学论证上无可挑剔，但是它所依据的公理则长期受到质疑。尽管如此，由于期望效用函数在理论上简洁易用，它在经济学研究中始终处于主导地位。而从认知心理学的角度来看待同样的问题，思路几乎完全不同。他们要考虑感知、信念、情绪、心态等许多方面，以至决策变为一个复杂的交替过程。这两位心理学家就是出于这样的考虑提出他们的所谓“小数定律”(人们根据少量经验就进行推理)、“展望理论”(不追求期望效用最大的一种决策过程的描述) 等等。不过，这并不是说他们的理论与期望效用函数理论完全对立，而是说前者代表人们在不确定环境下决策的完全理性行为；从长远来说，人们在实践中不断总结经验，其行为会越来越接近于这种理想化。后者则代表人们在复杂的现实条件下可能有的“非理性”行为，它可能在许多情况下更接近于人们的实际行为。这样的区别对于建立适用于长期稳定状况的理论框架来说，或许并不重要，但是对于瞬间万变的金融市场来说，则提供了一种说明短期异常的有力手段。所谓“行为金融学”就在卡尼曼－特浮尔斯基的研究的推动下，蓬蓬勃勃地发展起来。

从这两年的诺贝尔经济学奖的颁奖看来，人们在力求走出过于理想的一般均衡框架。考虑不对称信息、非理性行为都是其中的重要手段。它们都在一定条件下，能说明市场中的一些“反常现象”。然而，我们也可看到，这些新理论的提出，并没有“彻底摧毁”原有的一般均衡框架或者经典的金融经济学。事实上，直到现在为止，如果最终仍然要回答某个时期金融商品是如何定价的，那么某种稳定的均衡状态仍然是需要的。否则哪里还有什么相对稳定的价格或价值可定。区别仅在于市场中的经济活动者可能由于个人的信息、信念、心态、偏好等等的不同，使得结果与经典的讨论有较大的差异。既然如此，理论上应该还可能在一个更大的框架上把它们统一起来。

这里特别要提一下，柯克兰 (John H. Cochrane)于 2001 年出版的《资产定价 (Asset Pricing)》一书。这本专著正是企图在更高的层次上建立也适用于信息经济学和行为金融学的金融经济学的框架。虽然该书的基本理论结果仍然类同于罗斯的资产定价基本定理，并且许多理论推导也都已在以前问世，但是把它明确表达和总结为适用于金融学经典和各种新发展的形式，应该说是该专著的贡献。这里所说的统一框架是指文中提到的无套利假设的更确切的数学形式。尤其是金融经济学的研究发展已经发现，资产定价问题的答案虽然由于引入“信息”、

“行为”等有所变化，但是有一条基本法则没有改变，那就是线性定价法则，即一个 (未来价值不确定的) 资产组合的 (当前) 价值应该等于其组合成分的 (当前) 价值之和。所谓莫迪利阿尼－米勒定理其实说的就是这件事，它比完全的无套利要求要低一些，即在这一线性定价法则下，仍然有可能一个未来值钱 (价值为正) 的资产组合，当前可能不值钱 (价值非正)；也就是说，套利仍然可能存在。而更令人惊奇的是，可以证明，马科维茨证券组合理论和资本资产定价模型都是与线性定价法则等价的，即在一个金融资产市场上，如果有一条为金融资产定价的线性定价法则，那么它等价于市场上存在某条组合前沿，或者存在对某个无风险证券和某个“市场组合”，资本资产定价模型成立；反过来也一样，组合前沿的存在或者资本资产定价模型成立，也等价于某条线性定价法则的成立。不过，这里不包括布莱克－肖尔斯期权定价理论，后者仍然需要完整的无套利假设，或者按目前的术语来说，它服从的是某种“正线性定价法则”。

这样一来，“新金融学”与“经典金融学”之间的差别仅仅是线性定价法则的具体定价差别，或者说差别仅在于“基本金融资产”的“单价定价差别”，而在定价的数学形式上，它们都是一致的，并且经典金融学的许多结论仍然保持。这种“具体定价差别”表现为一个随机折现因子，即任何未来价值不确定的金融资产的当前价值等于其 (随机) 未来价值与随机折现因子的乘积的期望值。而不管是经典经济学，还是信息经济学、行为经济学对资产定价理论的应用，都归结为如何确定这个随机折现因子。理论的实证研究也同样如此。这样的新框架令人耳目一新。它至少使人不再对层出不穷的“非理性理论”感到无所适从。确实，至今我们还没有看到过针对“买十送一”之类的“非线性定价理论”。当然，正如该专著所指出，随机折现因子理论本身只是一个空洞的框架。其经济学内容仍然需要对经济学现实进行深入研究才能得到。

总之，在这 21 世纪初的前两年，这两届诺贝尔经济学奖和这本专著，给我们带来“信息”、“行为”和“随机折现因子”这三个关键词。看来我们应该密切注意这三个关键词对金融经济学发展的深远影响。

一、金融经济学的基本思想

本讲要求：了解金融经济学的基本思想和数学公理化的研究方法。对线性定价法则与无套利假设的运用有初步印象。

数学预备知识：初等代数。但 1.7 需要初等概率论和初等微分学。

内容：

- 1.1 金融经济学简史及其基本文献。
- 1.2 数学公理化方法及其有关争论。
- 1.3 作者的态度。
- 1.4 商 (管理) 学院学生为什么要学理论金融经济学。
- 1.5 从逻辑角度来看数学公理化方法的优势和缺陷，尤其是对金融经济学研究来说。
- 1.6 一个简单的模型说明线性定价法则和无套利假设怎样用来进行期权定价。
- 1.7 一个简单的投资－消费模型及其与无套利假设的关系。
- 1.8 有关课本和专著的介绍。
- 附录：一个说明“公共知识”概念的寓言。

我们这门课的题目叫《金融经济学》。由于在中文中“金融”两字的含义很广，这个题目可能会有各种不同的理解 (例如理解为货币银行学等)。但是我们这里的理解是最近半个多世纪中发展起来的理论金融经济学。这里的金融经济学主要指金融市场的经济学，与财政、财务、货币、银行、保险、税收等等，都没有多少直接的关系。理论金融经济学在这五十年间形成了一个完整的体系。我们的目标是要对这个体系有比较完整的了解。虽然以下的章节从名称上来看，大部分都是读者可能已经在《证券投资学》之类的课程中遇到过。但是在那些课程中，有关的内容更多地是作为实用方法来介绍。而我们的课程则要着重从理论的角度来考察，或者可以说，从数学公理化的角度来考察。

1.1　在“代引言”中我们已经回顾了从数理经济学到数理金融学的发展。数理经济学与理论经济学的内涵之间或许还有相当大的距离，但数理金融学与理论金融经济学则没有多大差别。这里我们首先要说，金融经济学的出发点还是新古典主义经济学。在它的历史上第一个要提到的还是 1954 年 Arrow (1921–) 和 Debreu (1921–) 的一般经济均衡理论。由此我们可以排出下列时间表：

- Arrow–Debreu 一般经济均衡存在定理
  Arrow, K., and G. Debreu, 1954, Existence of an equilibrium for a competitive economy, *Econometrica*, **22:** 265–290.

- Markowitz (1927–) 投资组合选择理论
  Markowitz, H., 1952, Portfolio selection, *Journal of Finance*, **7:** 77–91.
  Markowitz, H., 1959, 1991 Second ed., Portfolio Selection: Efficient Diversification of Investment, Basil Blackwell, Cambridge (有中译本：马科维兹，H.，1999，资产组合选择和资本市场的均值－方差分析，朱菁, 欧阳向军译，上海人民出版社，上海三联书店)。

- 关于公司理财的 Modigliani (1918–)–Miller (1923–2000) 定理 (Modigliani–Miller Theorem, MMT)
  Modigliani, F., and M. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review*, **48:** 261–297.

- Sharpe (1934–), Lintner (1918–), Mossin 等的资本资产定价模型 (Capital Asset Pricing Model, CAPM)
  Sharpe, W., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, **19:** 425–442.
  Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, **47:** 13–37.
  Mossin, J., 1966, Equilibrium in a capital asset market, *Econometrica*, **34:** 768–783.

- Fama (1939–) 市场有效性理论
  Fama, E. F., 1965, Random walks in stock market prices, *Financial Analysts Journal*,. September/October.
  Fama, E. F., 1970, Efficient Capital Markets: A review of theory and empirical work, *Journal of Finance*, **25**: 383–417.

- Black (1938–1995)–Scholes (1941–)–Merton (1944–) 期权定价理论
  Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, **81:** 637–654.
  Merton, R., 1973, The theory of rational option pricing, *Bell Journal of Economics and Management Science,* **4:** 141–183.
  Merton, R. C., 1992, Continuous-Time Finance, Rev. ed., Blackwell, Oxford.

- Ross (1944–) 套利定价理论 (Arbitrage Pricing Theory, APT)
  Ross, S. A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, **13:** 341–360.
  Ross, S. A., 1978, A simple approach to the valuation of risky streams, *Journal of Business*, **51:** 453–475.

至此，金融经济学的大厦就完全成形。上述这些经济学家，除了已故的 Black 以及 Lintner, Mossin, Fama 和 Ross 外，都先后获得了诺贝尔经济奖。这一大厦是有共同基础的整体。这门课希望能把这一整体进行概括的逻辑描述。

1.2 现代数理经济学是以 1954 年 Arrow–Debreu 一般经济均衡存在定理的出现为标志的。1959 年，Debreu 把他的学位论文以《价值理论》为标题正式出版，从此开创了数理经济学的一个新纪元。这一新纪元的特征在于它完全采用了数学公理化方法，把经济学的结论分为两类：一类是“公理”，它们是被假定成立的理论出发点；另一类是“定理”，它们是前者的逻辑推论。如果你承认“公理”是正确的，那么你必须承认“定理”也是正确的，除非你认为这个世界根本无逻辑可言。为使这种论证方法可行，首先需要对一个具体的理论问题建立数学模型。这个模型可以很简单，也可以很复杂。在那些复杂的金融数学模型中几乎涉及所有现代数学学科，尤其是微分方程，数理统计，随机分析，非线性分析等学科。但是作为初学者，我们也可通过简单的数学模型来理解其中的经济思想，并把它们用到各种金融问题中去。当然，如果真要对金融经济学深入研究，甚至仅仅是具体运用那些理论成果，艰深的数学工具是必不可少的。

这样的数学公理化方法对于经济学研究究竟起什么作用是一个长期引起争论的问题。这里争论的并非是数学对经济学有没有用。对此人们基本上已有共识：一些起码的数学工具 (正如这份教材所规定的起点：多元微积分，线性代数，初等概率论与数理统计等) 对于任何性质的经济学研究几乎都是必不可少的。问题在于经济学是否一定需要纯而又纯的数学公理化框架，以及由此引起一系列必须用专业数学家掌握的高深工具才能进行的研究。对于这个问题的正面回答是我们已经在“代引言”中引过的 Debreu 的话。这里我们不妨再重复一遍：“坚持数学严格性，使公理化已经不止一次地引导经济学家对新研究的问题有更深刻的理解，并使适合这些问题的数学技巧用得更好。这就为向新方向开拓，建立了一个可靠的基地。它使研究者从必须推敲前人工作的每一细节的桎梏中脱身出来。严格性无疑满足了许多当代经济学家的智力需要，因此，他们为了自身的原因而追求它，但是作为有效的思想工具，它也是理论的标志。”反对者则主要指责数学公理化

脱离实际。他们引用凯恩斯 (Keynes, J. M., 1883–1946) 1936 年所说的话：“目前过多的数理经济学只是一种大杂烩，和它所依赖的初始假设一样不精确，它使作者在矫揉造作的、无用的符号的迷宫中丧失了对现实世界的复杂性和相互关系的洞察力。”，并且指出：“所有这些根本不是把数学运用于现实的经济问题，相反，是把高度精确、复杂的数学运用于完全想象中的、幻想的经济学的美妙的理想国。”(见 [1], 149–151。)

1.3 对于这一问题的态度，很容易发现，参与争论的人们似乎很难摆脱由其自身知识结构所带来的 (正面的或负面的) 影响。本讲义作者原本是专业数学工作者，若干年前进入金融领域，自然很希望进行“把高度精确、复杂的数学运用于完全想象中的、幻想的经济学的美妙的理想国”那样的研究。这可以使你大可不必每天都要关注变化无常的股市行情，日新月异的金融变革，以及叫人心烦意乱的大千世界，而只需关在美妙的“理想国”中做自幼就热衷的数学题。但是如果我们要把回答金融现实提出的问题作为己任，或者至少要明白这个“理想国”并非完全是太虚幻境，那么就必须指出“理想国”的实在含义。这一问题显然不能简单地用“Debreu 等这样的研究者得了诺贝尔经济学奖”，或者“在美国一流经济学杂志上的许多论文都是那样的”来回答。

关于这样的问题的全面论述自然不是这份讲义的任务。事实上，这类问题看来也永远要争论下去，不可能在短期内使大家有共识。有兴趣的读者可参看若干年以前，北京大学出版社翻译出版的两本书 [1] 和 [2]，以及近年来经济科学出版社出版的 [3] 和 [4]。不管读者本人对于学习经济学持什么态度以及对书中的一些观点是否赞同，这两本书中所进行的讨论都是发人深思的。在这里，我们只想简单地表明作者的基本观点。

首先，我们赞同 Debreu 的说法。即，数学公理化方法是一种“有效的思想工具，它也是理论的标志”。这里不涉及对新古典主义经济学本身是否赞同的问题。其实有些对数理经济学的批评并非针对数学公理化方法，而更多的是针对新古典主义经济学本身。对此，我们应该把它们区分开来。为什么数学公理化方法是“有效的思想工具”，“理论的标志”？这里又涉及我们在这里不能详尽论述的什么是科学理论的问题。我们对这一问题谈一些粗浅的认识。我们认为，尽管人们对波普尔 (Popper, K. R. 1902–1994) 的批判理性主义有各种各样的看法，但是他对什么是科学理论的论述还是可以作为一种“理论的标志”来接受的。波普尔认为，科学理论是一种“全称陈述”，因而是一种“记号或符号的系统”。“系统必须表述得足够清楚和明确，使得我们易于辨认出每一个新假定是一种系统的修改。”因此，“一个严密的系统的形式被作为目的来追求”。“这种形式”就是所谓“公理化系统”。“公理或

者可以被看作是 (i) 约定，或者可以被看作是 (ii) 经验的或科学的假说。”“在一个理论系统内，我们可以区别属于各种普遍性水平的陈述。普遍性最高的陈述是公理；较高水平的经验陈述相对于从它们演绎出来的较低水平的陈述来说，总是具有假说的性质”(参看 [3])。而科学与伪科学、非科学的划界在于它的“可检验性”和“可证伪性”，其中最重要的是“通过能从理论推导出的结论的经验应用来检验理论”以及“经验的科学系统必须有可能被经验反驳”。波普尔在这里强调的是：科学研究的逻辑是演绎逻辑，不是归纳推理；而一个科学理论必须能经得起实践的检验，并且在检验中不断发展。在这一意义下，数理经济学正是经济学向科学化发展的一个标志。我们经常援引的据说是马克思 (Marx, K., 1818–1883) 的名言[1]：“一种科学只有成功地运用数学时，才算达到了真正完善的地步。”也可使人们在这一意义下得到进一步的理解。

当然，这并不意味着我们全盘接受波普尔哲学，而仅仅是在“理论的标志”这点上接受玻普尔所提出的要求。对此，我们在这里不得不提到 M. Friedman 于 1953 年 ([11]) 提出的观点。Friedman 认为：“把理论看作实质性假说的一个整体，理论应该通过对它试图 ‘解释’ 的现象类的预言能力来加以判断。只有实际证据才能表明它是 ‘正确的’ 还是 ‘错误的’，或更确切地说，是试验性地 ‘被接受’为有效，还是‘被拒绝’。”

这段话在精神上与波普尔相当一致，但 Friedman 说他当时对波普尔的主张并不熟悉。然而，由于 Friedman 认为对理论的假设前提的检验是不必要的，甚至声称假设前提的虚假不仅不是一个缺点，而且是实证经济学的一个优点。这就在经济学界引起了一场大论战。 Friedman 也被冠以“工具主义”，并且还被当作整个芝加哥学派的主要思想。对 Friedman 的工具主义不但有包括 Samuelson 在内的许多有影响的经济学家的批评，其漫画式的夸张，更是常常成为人们用来讽刺经济学家的笑料。[2]

对此，我们的第二点看法是许多假设前提很难检验的经济理论实际上是人们认识事物的一个阶段。Friedman 工具主义与其说是一种经济学方法论，不如说是当代经济学的相当大部分的无可奈何的现实。在人类的认识仅仅达到目前的水平时，工具主义的态度可能是对人们阶段性的认识的一种自我解嘲式的辩护。如果不把它们无限夸大，在一定程度上也是可接受的。就拿一般经济均衡理论作为主体的新古典

---

[1] 实际上，康德 (E. Kant, 1724–1804) 说过类似的话。

[2] 几乎每一个嘲笑经济学家的笑话都似乎与 Friedman 工具主义有关。例如，几个学者被困在一个荒岛上，他们有一个罐头无法打开。物理学家、化学家纷纷提出各种利用太阳能之类的物理、化学手段，而经济学家的解决方案是“让我们假设我们有一个开罐头的起子”。又如，经济学家跑到路灯底下来找他在一个黑屋子里丢失的东西。如此等等。

主义经济学来说，其假设前提都很难经得起实践的检验（你能验证你的日常消费遵循“效用最大化原则”吗？）。在上面提到的前两本书以及其他一些“非均衡”、“反均衡”之类的著作，对此都有许多一针见血的批评。但是直到现在为止，一般经济均衡框架依然挺立在经济学王国的土地上。而所有企图取代它的理论框架对这挺立了一个多世纪的越来越复杂的大厦来说还是相形见绌。这是很难仅仅用学术界的“惰性”来解释的。我们的观点是在我们的“代引言”中的一句可能不显眼的话：“直到现在为止，一般经济均衡理论仍然是唯一的对经济整体提出的理论。”也就是说，虽然已经有不少“反均衡”、“非均衡”之类的“经济理论”。但是如果以上面理解的“理论的标志”来衡量是否够得上“科学理论”，那么“一般经济均衡理论仍然是唯一的对经济整体提出的理论”。其他“理论”都未达到那样的高度。尽管一般经济均衡理论在“可检验”、“可证伪”上并不令人十分满意，但是基本上还是相容的。至于金融经济学中的 Friedman 工具主义似乎比别的经济学分支中更为浓重。Sharpe 在其有关 CAPM 的经典论文中，就明确提出他的研究的假定是“高度受限制的 (high restrictive)”和“无疑是不现实的 (undoubtedly unrealisitic)”，但其逻辑推论是“可接受的”。这就是一种 Friedman 式的态度。Black–Scholes–Merton 理论的建立也同样不是其假设前提的可检验性，而是理论本身的逻辑结构及其广泛的可应用性。

第三，我们愿意在一定程度上接受“多元主义”的认识论。如果我们以认识在多大程度上“合乎逻辑”来作为感性认识和理性认识的衡量标准，那么除了极为简单的事物以及力学或其他自然科学的某些分支以外，各种感性－理性程度的认识对于人们把握事物、作出判断都是有益的。它们都在某个侧面反映了现实世界。这里极端的“感性认识”以仰望天上的白云，引起无限遐想为例；而极端的“理性认识”则是“理想国”数学公理化体系中的推理。人类的文化也许也可这样从极端的“文学文化” (浮想联翩的意识流小说？) 到极端的“科学文化” (只见符号不见文字的数理逻辑体系？) 来进行归类。当然，我们这里并不是要抹杀现象与本质的区别。但是对于像经济、金融那样复杂的系统，人们的认识能力还不足以用足够完整的反映本质的“理论体系”来把握它时，那么必须通过搜集、列举大量现象“案例”来进行补充。界于它们之间的是福尔摩斯式的 (不完整) 推理、统计分析等等。我们不应该像那几个各摸着大象的一部分的瞎子那样，声称只有自己摸到的那部分才是大象，而是应该把各自摸到的一部分互相补充，以综合成一个较为完整的大象形象。一般来说，偏于“文学文化”的思维方式往往整体外观较强，但较少考虑内在的逻辑联系；而偏于“科学文化”的思维方式则有时会相反。遗憾的是，由于在学科领域上、在社会活动范围上的局限，不同文化圈的学术工作者经常会互相排斥。在这里，也许可以说，商 (管理) 学院是一个很特殊的地方，它似乎能够成为一个各

种“感性－理性程度”的文化的大熔炉。商 (管理) 学院的学生应该接受各种类型的文化熏陶。

1.4　以上就是我们的基本态度。其实归根结底，我们想说的是为什么商 (管理) 学院的学生需要学习理论金融经济学。对于金融的认识或许特别适用上面提到的第三点。金融领域是一个对我们的世界和社会几乎起主导作用的领域，尤其是在当今这个金融全球化、电子化、网络化的时代。但是对金融的认识，我们不仅是通过财务管理、证券投资等这样的课程，也通过茅盾的小说《子夜》(或者更现代的梁凤仪的小说)、奥立弗·斯通的电影《华尔街》；我们既要读 W. Sharpe, R. Merton 等这些诺贝尔经济奖得主的投资学教科书和专著论文，也不得不注意索罗斯、巴菲特这些自成一派的投资家的经验之谈或者什么“金融炼金术”；既要熟悉股市的走势图、K 线图、点线图、各种技术分析指标，以至道理论、甘理论、艾略特波浪理论等等，又要对分形几何、数据挖掘、神经网络、遗传算法等各种股市分析预测方法感兴趣。甚至连星象算命、易经八卦也常常挤进金融市场来争他们的一席之地，更不要说我们要经常注意格林斯潘在说什么，因为他的一番谈话有时会让全球金融市场发抖。这些人们从各个角度对金融领域的认识，除了完全虚构的算命八卦之类，很难说有哪方面完全没有意义。而精通各种技术的基金公司常常使基金收益跑不赢无意识的大市似乎是是常事，由诺贝尔经济奖得主 Merton, Scholes 等人发起的长期资本管理公司 (Long-Term Capital Management) 在这两位获得诺贝尔奖后的第二年 (1998) 差点破产，又成了轰动一时的丑闻。这些又给人造成一种错觉，以为金融学不可能有科学理论，而只需要对市场察颜观色的经验。这种错觉当然离真理更远。

我们的结论是：理论金融经济学是代表人类对金融领域的最为“理性”的认识。它赋予我们对金融现实进行理性思考的基本框架。因此，为了深入揭示金融现实的本质，这样的理论框架是绝对必要的。但是由于金融现实的无比复杂，这样的理论框架直到半个世纪以前才开始逐步成形，并且在这一框架中能揭示的金融的本质仍然相当有限。因此，人们还是需要通过许多其他的手段来认识金融。作为一名商 (管理) 学院的学生，人们可以根据自己的知识结构上的优势，来为自己设计认识金融的途径。他们不一定要成为金融经济学理论的研究工作者，但是在他们的知识结构中，如果对理论金融经济学一无所知，无论从对金融现实的理解，还是从与金融理论研究人员交流的角度来看，都将是一种很大的缺陷。

这里我们还要指出的是，金融经济学理论其实比它的先驱框架一般经济均衡理论要幸运得多。说一般经济均衡理论是空中楼阁，就像前面提到的两本书 [1], [2] 中所说到的那样，似乎还真是那么一回事。

但是要说金融经济学理论是空中楼阁，那是站不住脚的。Markowitz 组合选择理论在基金管理中是常规武器；资本资产定价模型 (CAPM) 的检验虽然有很大争论，但是无论是理论还是实际应用， CAPM 还是无处不在； Modigliani–Miller 定理尽管也显得有点脱离实际，但是毫无疑问，它奠定了公司财务这一学科的基础；Ross 的套利定价理论有点牵强附会，但不失为一种非常有用的工具；至于 Black–Scholes 期权定价公式与现实吻合得如此之好，不但是它最终问世的主要原因之一，更由于它久经考验，使人们对它的成立已经深信不疑。相反，人们还常常认为，当 Black–Scholes 公式不成立时，并非是它的过错，而是“市场出了错”。事实上，正是 Black–Scholes 理论推动了衍生证券市场的迅速发展，并促使无数金融机构聘用数学、物理博士去为他们计算衍生证券、对冲策略的价值。即使信息经济学、行为金融学等方面的研究正在不断用“反常现象”的实证分析来冲击经典金融经济学的一些基本论断，金融经济学的基本框架也仍然没有动摇。在这样的情况下，人们不但应该把金融经济学的研究成果作为一些分析技术方法来掌握，更应该寻根问底地对它们问个为什么，以求更好地发挥这些工具的作用。

1.5　不过话又要说回来。今天的金融经济学理论对于认识金融现实仍然是远远不够的。对于目前金融理论研究的一些热点，短期内对它们的研究深度似乎还很难够得上我们前面所提到的“理论的标志”。我们这里所指的是涉及所谓行为金融学、信息经济学范畴的一些问题。关于这些问题，我们可以从逻辑的角度来考察。

我们首先要注意到，波普尔对科学理论的第一个要求是它必须用**演绎逻辑**，而不是**归纳推理**，因为归纳推理是靠不住的。所谓归纳推理就是根据以往的经验，推断出一般的规律。例如，当人们看到 10 个乌鸦都是黑的，就断定“天下乌鸦一般黑”。但是波普尔指出，这样的推理导致逻辑矛盾是早在休谟 (Hume, D., 1711–1776) 的著作中已经说得很清楚。事实上，根据过去的经验而作判断的理由，是因为根据过去的过去的经验，曾对过去作过正确判断。于是这样不断去寻求“理由”，就会发现最后什么理由也没有找到。一切都建筑在不可靠的“过去曾发生过的事将来看来也要发生。”波普尔把演绎逻辑作为科学理论的要求就已经在学术领域中排除了许多学科的“理论”。尽管如此，这并不是说人们应该排除归纳推理来作为认知手段。事实上，每个人在日常生活中用得最多的是归纳推理。有时还带有浓厚的感情色彩 (例如，“这个人看来和善可亲”等等)。在金融学领域中，案例调查、统计分析、计量经济学模型等等，所使用的都是归纳推理。更不用说，股评家们所使用的“大盘将出现冲高震荡走势，短线操作应以观望为主”之类的语言。对于使用归纳推理的见解不能说一定没有价值。但是从逻

辑的角度 (理论依据？) 来看，它们的可靠程度是不一样的。

其次，如果把演绎逻辑就归结为数学公理化方法，那么这种逻辑至少对于金融经济学来说是远远不够的。直到现在为止，这种方法 (至少是人们通常使用的方法) 基本上是所谓“一阶谓词演算”，或者简单地说，是一种**一阶逻辑**。它考虑的是一些逻辑客体之间的同一层次的相互关系 (即所谓“谓词”)。而不太考虑，或者说还不太清楚怎样来考虑，不同层次的相互关系。后一种逻辑就是所谓**高阶逻辑**。这里我们举例来说明高阶逻辑意味着什么。其实最简单的二人对策问题中就已包含高阶逻辑问题。两人下棋，棋手通常并不是只考虑目前的局势下，最好的一着是什么，而是还要考虑对手是怎么想的，以及对手是怎么想我怎么想的，对手以为我这样想又是怎么想的。如此等等。这样的逻辑就不是我们通常运用的数学公理化方法的逻辑。目前在对策论中常规的处理方法常常是把它归结为一阶逻辑力所能及的问题来考虑，例如，只考虑有限种情形，利用极小极大策略等等。另外的方法还有 Bayes 统计推断等。但是我们会感到这样的做法实际上是很笨拙的。这是因为我们会用的数学和逻辑工具只有这些。希望更切题的工具还有待数学和高阶逻辑演算的发展。从金融经济学来看，正是由于工具上的限制，使我们对许多重要问题的研究还显得力不从心。最司空见惯的“炒股心理学”问题，正如凯恩斯所说的那样，如同一个“选美”问题 ([6]，133–135)：你不是要选一个你认为最美的人，而是要选一个大家认为最美的人。它就是一个高阶逻辑推理问题。著名的 Grossman–Stiglitz 悖论 (1976) 的粗浅说法是：如果市场价格已经反映了所有信息，那么投资者为什么还要去搜集信息？但是如果所有投资者都不去搜集信息，那么市场价格又怎么可能反映所有信息？ (参看 [7]) 其实也是因为涉及高阶逻辑问题。一个私下谁都知道的消息与它被正式颁布的作用是很不一样的。附录中的一个寓言性的故事相当充分地说明了这一点。于是“公共知识 (common knowledge)”在信息经济学和金融经济学中成为很重要的概念。但是传统的数学方法还不太会处理这样的概念。这使得目前很热门的行为金融学、金融市场微观结构理论之类的金融经济学分支虽然也用点数学，却很难给出全面的公理化框架，而只能满足于简单的代数方程、图表分析等等。因此，不要以为数学在金融中已经用得太多，实际上，数学的发展还大大赶不上金融的需要。

最后，还要注意的是数学公理化方法用的是**外延逻辑**，而不是**内涵逻辑**。所谓外延逻辑是指它所涉及的对象和集合都是由它们的外延 (即由它们的成员) 来确定的。两个集合由同样的成员所组成，就认为它们之间没有区别。一个集合中的两个元素被认为它们有同样的性质。它们之间可以互相替代。同一个元素在任何场合都可以替代它自身。但是内涵逻辑则要对这些根据对象规定的内涵，根据不同的场合，加以区别。例如，在证券选择理论中，证券被抽象为仅仅用一个代表证

券收益率的随机变量来刻划。如果两种证券的收益率完全一样，那么在理论中就导致它们没有区别。但是如果把这样的研究结果直接去应用时，那就会发现，在很多情况下是行不通的。现实中的两种证券，即使它们的收益率几乎完全一样，一般也是不能互相替代的，因为这两种证券还有许多不同的内涵 (基本面分析) 要考虑。投资策略与投资组合在内涵上当然相当不同，但是在外延上，由于它们都对应同样的量，经常被认为是同样的集合。期权定价理论中的基本方法是用基本证券的组合来“复制”期权。于是在外延上，期权与证券组合是可以互相替代的。但是在内涵上，这两者绝对不是一回事。以前遇到这样的问题时，我们常常简单地归结为理论与实际的矛盾，或是抽象与具体的矛盾。于是有时会以为通过把模型做得更细，增加更多的参数，情况会有所改善，很少想到这里其实有逻辑上的根本困难。内涵逻辑的研究目前尚处于初步阶段，除了少数专家以外，极大多数的研究者都对它很陌生。但是对于行为金融学、信息经济学研究来说，这将是一种人们期待的逻辑工具。而与它相比，外延逻辑的局限性也是显而易见的。

我们提出这些问题是希望引起读者注意。在我们的金融经济学课程中将不对这些问题进一步展开。这也是一种常规的做法。但这并不说明所涉及的问题不重要，相反，在有些金融问题的分析中，例如，股市中的做庄现象、跟风现象的理论研究，它们其实处于核心地位。

1.6　现在我们再回到我们给自己划定的范围中来。我们说，现代理论金融经济学研究的中心问题是金融资产的定价。Arrow–Debreu 模型回答了普通商品的定价问题。这一模型假定消费者追求最大消费效用，生产者追求最大生产利润 (它们就是作为理论出发点的“公理”，与几何公理的“天经地义”的明显不同在于，这些“公理”都是大可争辩的)，然后在一定条件下，存在一个一般经济均衡的价格体系，使得商品的供需达到均衡。金融资产的定价似乎也应该走这条路。但是由于金融市场的最主要的特征在于未来的不确定性，沿“均衡定价论”的道路前进步履十分艰难。1958 年 Modigliani 和 Miller 开始提出无套利假设作为“公理”来作为金融资产定价的出发点。这条“公理”其实只是“均衡定价论”的推论，即达到一般均衡的价格体系一定是无套利的。但是把它单独列出来，可以脱离“均衡定价论”的复杂框架，直接对金融资产定价。于是对于金融经济学来说，又逐渐发展成为一套“套利定价论”。这一理论十分有效。 Black–Scholes–Merton 理论几乎完全基于此。Ross 套利定价理论当然也是如此。不过“套利定价论”只能就事论事，由此建立不起来全市场的理论框架。因此，它只能作为“均衡定价论”的补充。但是正如“代引言”中引述的 Miller 的观点，对商 (管理) 学院学生所关心的“微观规范”金融学来说，“套利定价论”可能显得更为重要。

无套利假设类似于普通商品定价问题中的“无投入就无产出”假设。由于在金融市场中最后都会以钱来结算，所以投入和产出都将是钱。所谓无套利假设就是“无钱投入就无钱产出”。这就是现代理论金融经济学中的一条“公理”。这条“公理”显然只在非常理想的市场条件下才会成立。尤其是它被证券市场中的“技术分析派”所排斥，因为后者的出发点就是不断在证券市场中寻求套利机会。不过无套利假设是否成立可看作市场是否有效的标志。套利机会很多的市场显然是不稳定的市场，或者说市场不够有效的市场。在这样的市场中定价问题不可能有解。而定价问题可以作为理论问题来研究时，必须要求市场有这样的无套利假设。

现在我们就来看看无套利假设怎样用来给金融资产定价。如上所述，光有这样一条“公理”还不够，还需要对一个具体问题建立数学模型使“公理”具体化，才能进行推理。我们这里讨论的问题就是期权定价问题。所谓 (股票买入) 期权是指以某一固定的执行价格在一定的期限内买入某种股票的权利。如果执行期是固定的，则称为欧式期权；如果执行期可以是到期前的任何时候，则称为美式期权。期权在它被执行时的价格很清楚，即：如果股票的市价高于期权规定的执行价格，那么期权的价格就是市价与执行价格之差；如果股票的市价低于期权规定的执行价格，那么期权是无用的，其价格为零。不过，如果股票在未来到期时的价格从当前来看时是不确定的时，期权到期执行时的价格也是不确定的。现在要问期权在其被执行前应该怎样定价？为此，我们来考虑一个最简单的股市数学模型。这个模型中一共只有“当前”与“未来”两个时刻，并且“未来”时刻就是期权的执行时刻。“当前”的股价 S_0 是已知的，而“未来”的股价可能有两种情况，或是变成 aS_0, 或是变成 bS_0, 其中 $a \neq b$, 否则未来的股价就确定了。为简单起见，我们还假定“当前”与“未来”的货币价值是一样的，既无通货膨胀，也无银行利息。

现在假定期权的执行价格为 K. 那么期权在“未来”可能有的两种价格为 $C_a = \max(aS_0 - K, 0)$ 和 $C_b = \max(bS_0 - K, 0)$, 其中 max 表示括号内两个数中的大者。根据这样的一些假定，我们来求期权的“当前”价格 (下面的讨论实际上与 C_a 和 C_b 是什么无关，即它适用于任何类似的衍生证券的定价)。

在这里，我们首先要注意到，“期权可能定价”这件事本身就已经体现了最低要求的无套利假设。否则，如果对于未来有不确定价值的期权来说，其当前价值也不确定，那么它在市场上就不可能交易；或者说，在市场上同一种期权可能有多种价格，从而就有人可以进行“低买高卖”的套利活动 (这里我们不考虑买卖差价、交易费用之类的问题)。换句话说，我们提出期权定价问题，已经默认了“每一种 (未来价值不

确定的) 期权都有其 (当前确定的) 价格”这一公理。其次，我们在下面的讨论中，还把“线性定价法则”作为公理。所谓线性定价法则，是指若干份 A 证券与若干份 B 证券在一起的证券组合的总价值，应该等于 A 证券价格的同样倍数与 B 证券价格的同样倍数之和。换句话说，也就是一个证券组合的价值应该等于它的组成证券的价值之和。这条“线性定价法则”如果不成立也意味着存在某种套利机会，即我们有可能利用“合起来”买卖一个证券组合与“分开来”买卖一个证券组合的差价，来构造一个赢得套利的交易策略。这里还蕴含着一个假定：对任何一种证券，人们都可自由地买卖。即使你不拥有这种证券，也可以把它卖出。也就是说，允许人们对证券进行“卖空”。

在这样的基本“公理”的假定下，期权定价问题就可通过几种表面上不同的观点来得到解决。其中主要的三种是：

1. 由于卖出股票与买入期权是两种风险方向相反的投资行为，适当组合这两种投资行为，就可达到完全保值的作用。这样，设 C_0 为“当前”的期权价格，投资者卖出一份股票，买进 x 份期权，而这两种投资行为的组合使风险完全对冲、价值完整确保，即不管出现哪一种情况，它的价值始终不变。于是由“线性定价法则”，这就得到两个方程：

$$-S_0 + xC_0 = -aS_0 + xC_a, \qquad -S_0 + xC_0 = -bS_0 + xC_b.$$

由此可解得期权的“当前”价格为：

$$C_0 = \frac{C_a - C_b + aC_b - bC_a}{a-b}.$$

2. 期权的“未来”价值虽然是不确定的，但是它完全依赖于股票的不确定的价格，以至期权本身可以通过股票交易和银行存款 (我们假定利率为零，有利息情形类似考虑) 的组合来“复制”。这就是说，一份期权相当于 y 元银行存款与 z 份股票的组合。于是由这种组合的“未来”价值与期权价值一致，再由“线性定价法则”就得到两个方程：

$$y + zaS_0 = C_a, \qquad y + zbS_0 = C_b.$$

由此可解得

$$z = \frac{C_a - C_b}{S_0(a-b)}, \quad y = \frac{aC_b - bC_a}{a-b}, \quad C_0 = y + zS_0 = \frac{C_a - C_b + aC_b - bC_a}{a-b}.$$

3. 我们只假定“未来”可能有两种情况，但并未规定这两种情况的可能性 (概率) 各有多大。每个投资者都可根据自己所掌握的信息对这两种可能作出自己的估计 (即所谓主观概率)。但是在进一步的无套利

假设下，就要导得 a 与 b 中必然有一个大于 1, 另一个小于 1; 即股市的两种情况只能是一涨一跌。否则在都上涨时，投资者可通过“当前”买进，“未来”卖出，稳能得利；在都下跌时，投资者可采用相反的策略，同样稳能得利。当然，我们这里同样还要假定投资者总有一定的资金可支配，并且股市允许“卖空”，即允许卖出你并不拥有的股票，只要你能在“未来”交割时，有资金到市场去把股票买回。这里的进一步的无套利假设是指：当假定“当前”与“未来”的货币价值一样时，不存在未来价值总高于当前价值的证券组合。

在这种假定下，就存在有一种对未来的可能性估计，使得“未来”的股价的平均值恰好就等于“当前”的股价。这是因为 a 和 b 中必然有一个大于 1, 另一个小于 1, 从而存在 0 与 1 之间的数 q (概率) 使得

$$aq + (1-q)b = 1, \qquad \text{即} \qquad q = \frac{1-b}{a-b}.$$

而“当前”的期权价格应该就是在这种可能性估计下的“未来”的期权价格的平均值，即

$$C_0 = qC_a + (1-q)C_b = \frac{C_a - C_b + aC_b - bC_a}{a-b}.$$

Black–Scholes 模型当然没有那么简单，其中最主要的一点在于他们假设时间的变化是连续的，股价的变化也是连续的。这对于允许投资者每天都可做任意多次交易的美国股市来说，是一个很好的刻画。这时股价变化被假定为在数学上用所谓几何 Brown 运动来表示；同时，还假定有一种其价值作指数增长的无风险证券 (债券或银行存款)，来作为股价的参照物。Black–Scholes 的原始论文中用的是这里的第一种观点，即股票交易与期权交易的适当组合可对冲风险，从而成为一种无风险证券。由此导得一个偏微分方程，其解就是 Black–Scholes 期权公式。后来的研究，尤其是 Merton 等的研究，又形成了另两种观点，它们在对 Black–Scholes 理论的深化和推广中，起了关键作用。第二种观点隐含着一种“完全市场”的概念。在完全市场中，每一种价值取决于股价变化的衍生证券 (理论上又称“未定权益 (contingent claim)”) 都可以通过某种股市交易策略来“复制”，从而它们的价格也就由股价和交易策略根据线性定价法则来决定。完全市场的概念在理论上极为重要。如果意外情况非常频繁，而股票之类的证券种类又不是足够多，那么相当多的衍生证券是不可能“复制”的，从而它们就不可能有唯一的定价。现实的市场总是不完全的，因而许多本质上不能定价的衍生证券交易必然包含本质的不可捉摸的风险。第三种观点表达为严格的数学形式后，称为“资产定价基本定理”。用不太精确的日常用语，它可以说成：(完整的) 无套利假设等价于存在对未来的不确定性的一种估

计 (数学上称为“等价鞅测度”)，使得任何时候的股价都等于未来股价的平均值 (数学上称为股价形成“鞅过程”；如果利率不为零，这里还要考虑股价的折现值)。这在某种意义上可以说，当市场充分发展到任何人都不能坐享其成时，未来的不确定性有其内在的规律，人们不应凭借一种赌博心理去胡乱臆测，而应力求去掌握这种规律。正是在这一点上，Black–Scholes 理论显示了它对股市实际以至更一般的不确定环境下的决策问题研究上的重要性。

不过，正如我们在上面已经注意到的，这里的前两种观点只要求“线性定价法则”，而第三种观点需要完整的无套利假设，否则 p 可能走出 0 与 1 之间，而不再有概率意义。尽管我们得到的最后结果在形式上都是一样的，但在推导中，第三种观点能保证当期权的未来价值非负时，其当前价值也一定非负。它其实也就是完整的无套利假设的要求。而在前两种观点下，并不能保证这一点。也就是说，“线性定价法则”并不能保证未来价值非负的期权，其当前价值也非负。事实上，我们很容易构造这种反例：例如，取 b 和 $a-b$ 足够小，$C_a=0, C_b>0$, 就能使 C_0 取任意的负值。尽管金融市场的实际使得通常不会出现这样的情况，但是这样的例子说明，“线性定价法则”与“存在套利”是不矛盾的，它也可能适用于金融市场中的许多“反常情形”。

从这一例子中我们可以看出现代金融经济学的推理方式。一般来说，其中用到的数学都比这里所说的要复杂得多。但是在基本思想上大致就是如此。对于数学基础较好的读者，可以沿着这样的思路走得很远。而对于数学上有一定困难的读者来说，通过这样的简单模型也能掌握理论的内涵，并且由此既可看到理论的威力，也可看到理论的局限性。当然，简单的模型没有实用价值。我们不能指望这一节中提出的模型就能用来进行期权定价的实际计算。

1.7　下面我们再对无套利假设和资产定价基本定理从另外一个角度来进行讨论。我们已经看到，通过线性定价法则或无套利假设，可以用基本证券来为衍生证券定价。这样的定价方法曾经长期为人们所不理解。原因在于传统的定价理论都是根据经济活动者的需求和供给，在一般经济均衡框架中形成的。而在上面的定价方法中，只不过是通过股票的价格和货币的绝对保值来为期权“相对定价”。它并不关心股票价格本身怎样形成这一问题。这里我们完全看不到市场中的经济活动者。这一矛盾在后续的研究中慢慢也搞清楚了。在这一小节中，我们再来构造一个十分简单的投资－消费模型，以说明从这个角度来看资产定价基本定理，它也是合理的。

我们的出发点仍然与前面一样：假设只有“当前”与“未来”两个时刻。“当前”的股价 S_0 是已知的，而“未来”的股价可能有两种情况，

或是变成 aS_0, 或是变成 bS_0, 其中 $a \neq b$. “当前”与“未来”的货币价值是一样的，既无通货膨胀，也无银行利息。假设有一名投资－消费者 A. 他所追求的目标是他的消费效用最大。为此需要定义他的效用函数。首先是有这样一个静态的效用函数 $u(c)$, 其中 c 是他所消费的价值。一个合理的假定为 u 是 c 的递增函数，即对 A 来说，“钱花得越多越好”。但是我们目前有“当前”和“未来”两个时刻，对 A 来说，还需要通过当前的确定消费和将来的不确定消费，来构成一个综合的效用函数。

在我们前面的讨论中，我们始终没有提及“未来”可能发生的两种状态的概率。事实上，我们在最后也看到，衍生证券的定价与股票的两种状态的发生的概率没有关系。但是在这里，当 A 要对自己的投资消费行为进行决策时，就不得不根据他所掌握的信息，对“未来”有一个估计。于是他估计股票价值为 aS_0 (以后我们称这一状态为状态 a) 的概率为 p, 而为 bS_0 (以后我们称这一状态为状态 b) 的概率就是 $1-p$. 这个 $(p, 1-p)$ 可能是确实要发生的“客观概率”，也可能是 A 根据自己掌握的信息来判断的“主观概率”。根据所谓 von Neumann–Morgenstein 的期望效用函数理论 (见第六讲)，A 的未来不确定消费的效用就等于其不确定效用的数学期望。也就是说，如果 A 在状态 a 的消费价值为 c_a, 在状态 b 的消费价值为 c_b, 那么其未来消费的效用就是 $pu(c_a)+(1-p)u(c_b)$, 这里我们假定未来的消费效用函数与当前消费效用函数是一样的。如果我们再假定，A 的总消费效用函数就是他的当前消费效用与未来消费效用之和，那么他的总效用函数就是

$$U(c_0; c_a, c_b) = u(c_0) + pu(c_a) + (1-p)u(c_b),$$

其中 c_0 是当前消费价值，c_a 和 c_b 分别是两种未来消费价值。这里实际上还假定 A 对当前和未来的重视程度一样。如果他重当前，轻未来，在这个效用函数的后两项前乘一个 0 与 1 之间的“折现因子”。但这并不影响下面的讨论。

现在我们要讨论的问题是：A 怎样通过他对银行和股市的投资，来使他的总消费效用最大。这里还需假定 A 在当前有资金 e_0, 在未来有资金 e_a 和 e_b. 于是如果 A 在当前向银行存款 x_d (如果 $x_d < 0$ 意味着贷款), 在股市买入股票 x_s (如果 $x_s < 0$ 意味着“卖空”)，在未来从银行取款，向股市卖出股票，那么我们就有

$$c_0 = e_0 - x_d - S_0 x_s; \quad c_a = e_a + x_d + aS_0 x_s; \quad c_b = e_b + x_d + bS_0 x_s.$$

这里我们仍然假设银行是无利息的。这样，对 A 来说的最优投资－消

费决策问题就是

$$\begin{cases} \max & U(c_0; c_a, c_b) = u(c_0) + pu(c_a) + (1-p)u(c_b) \\ \text{s.t.} & c_0 = e_0 - x_d - S_0 x_s \geqslant 0 \\ & c_a = e_a + x_d + aS_0 x_s \geqslant 0 \\ & c_b = e_b + x_d + bS_0 x_s \geqslant 0 \end{cases}$$

这一个问题是不是一定有解？不一定。事实上，只要假设 u 是递增函数，当 a 和 b 不在 1 两边，那么这个问题一定没有解。这是因为这时就会发生我们上面谈到的套利机会。A 只需要通过当前向银行借款，低价买入股票，未来高价卖出股票 (或者当前高价卖空股票，把钱存入银行，未来低价买入股票来交割) 就能增加他的总效用，并且这样做是无止境的。用一句数学的话来说，我们可以在变量的约束范围内使 u 趋向于无限大。另一方面，如果这一问题有解 $\bar{x}_d, \bar{x}_s$，那么作为上述结论的逆否命题，(完整的) 无套利假设一定成立。这就是说，如果我们需要用一般均衡框架来讨论金融问题，无套利假设的成立其实也是先决条件。这就进一步肯定了无套利假设在理论上的重要性。不但如此，如果在数学上假定一定的函数光滑性条件，那么我们还可从解的一阶必要条件得到

$$\left.\frac{\partial U(e_0 - x_d - S_0 x_s; e_a + x_d + aS_0 x_s; e_b + x_d + bS_0 x_s)}{\partial x_d}\right|_{x_d = \bar{x}_d; x_s = \bar{x}_s} = 0;$$

$$\left.\frac{\partial U(e_0 - x_d - S_0 x_s; e_a + x_d + aS_0 x_s; e_b + x_d + bS_0 x_s)}{\partial x_s}\right|_{x_d = \bar{x}_d; x_s = \bar{x}_s} = 0.$$

即

$$\begin{aligned} u'(\bar{c}_0) &= pu'(\bar{c}_a) + (1-p)u'(\bar{c}_b) \\ S_0 u'(\bar{c}_0) &= aS_0 pu'(\bar{c}_a) + bS_0(1-p)u'(\bar{c}_b) \end{aligned}$$

其中[3]

$$\begin{aligned} \bar{c}_0 &= e_0 - \bar{x}_d - S_0 \bar{x}_s \geqslant 0 \\ \bar{c}_a &= e_a + \bar{x}_d + aS_0 \bar{x}_s \geqslant 0 \\ \bar{c}_b &= e_b + \bar{x}_d + bS_0 \bar{x}_s \geqslant 0 \end{aligned}$$

由此可以得到，

$$\begin{aligned} 1 &= p\frac{u'(\bar{c}_a)}{u'(\bar{c}_0)} + (1-p)\frac{u'(\bar{c}_b)}{u'(\bar{c}_0)} \\ 1 &= ap\frac{u'(\bar{c}_a)}{u'(\bar{c}_0)} + b(1-p)\frac{u'(\bar{c}_b)}{u'(\bar{c}_0)} \end{aligned}$$

---

[3] 严格地说，以下的 $\geqslant$ 都应改为 $>$. 这在对效用函数 u 作更多的假定时是有保证的。如果没有严格的不等号成立，一般来说，最优解的一阶必要条件不一定成立。

与上面的第三种观点相比较，令 $q = p\frac{u'(\bar{c}_a)}{u'(\bar{c}_0)}$, 那么 $1 - q = (1 - p)\frac{u'(\bar{c}_b)}{u'(\bar{c}_0)}$. 这就得到了前面的“等价概率鞅测度”(请注意，由于 u 是递增函数，其导数 (“边际效用”) 总是正的)。也就是说，这其实就是另一种特殊方式的资产定价基本定理的证明。后一等式在假定未来股价 aS_0 和 bS_0 已知的条件下，给出了当前股价的定价公式。

上面的讨论说明，如果我们要从经济活动者的求最优投资－消费的问题出发，在得到股价定价关系的同事，我们同样可以得到资产定价基本定理。在这里它可表达为：**投资－消费最优问题有解的充分必要条件是无套利假设成立，或者资产定价基本定理成立**。这样，从无套利假设出发来讨论金融资产的定价问题，就有了更多的理论根据。

不过从上面的讨论中，我们也可看到一般经济均衡框架是一个多么沉重的十字架。首先是效用最大化问题就相当罗嗦，即使是对于我们这个简单到不能再简单的情形都是如此。我们需要对效用函数预先给出一系列相当人为的假设，然后又需求解比较复杂的数学问题。其次是对一个投资－消费者的问题求解，我们就可得到一个对于他来说的定价法则。但是全体投资－消费者的定价法则如何在市场上统一，又涉及一个理论上非常复杂的问题。因此，这也许只对研究“宏观规范”的经济学家有重要意义，而对只想知道“微观规范”的商 (管理) 学院学生来说，可能并不那么重要。

1.8 我们的讲义将这样来安排：先讨论从线性定价法则出发的“套利定价论”，再讨论“均衡定价论”。上面提到的这些经典工作将都会得到介绍。我们着重的是其理论实质，因而特别重视其基本思想和逻辑推理，但在数学上则不一定拘泥许多细节。适当的时候也会介绍一些应用实例，但不是这门课的重点。

金融经济学方面的教材、专著很多。常用的教材见文献 [10]–[11], [13]–[16]。被华尔街从业者当作“Bible (圣经) ”的是 Hull [16]，它已被再版多次。这本教材中实用和理论并重，但主要论述衍生证券。[13], [15], [16] 是 80 年代末出现的理论金融经济学的标准教材。可以作为进一步阅读的材料。但它们之间的差别非常大。在这些教材出现以前的标准教材是 Fama 和 Miller 于 1972 年出版的《金融理论(The Theory of Finance)》，以及 Fama 于 1976 年出版的《金融学基础 (Foundations of Finance)》。但是它们由于都不涉及衍生证券，已经赶不上形势的需要。讲义的思路比较接近于 Jarrow 的书。但是大大减轻可能使读者感到困难的数学形式化程度。 Duffie 的两本书 [10] 和 [11] 影响很大。它们在数学上要求更高些，适合于数学基础较好的读者。[9] 对于商 (管理) 学院的学生来说，值得一读。它告诉你 Black–Scholes–Merton 理论如何改变了传统的项目投资理论。Cochrane [8] 是最近出版的一本新著，它从

资产定价的角度来重新考察各种金融经济学问题。观点上与我们很相近，并且它对理论框架和实证方法的讨论几乎并重。这对金融经济学的实际应用更加有利。但它更多地是从投资－消费模型出发来建立其理论框架的。1.7 是该书基本思想的一个通俗化引言。至于强调“线性定价法则”(该书中称为“单一价格法则 (law of one price)”，这个名词可能由于 Pliska [19] 而得到传播，后者也是一本影响相当大的数理金融学的教材) 的作用，则是本讲义受到 [8] 的影响。这一影响使本讲义在前几稿的基础上作出大幅度的修改。

说明　在本章中，作者所表达的关于数学与经济、金融之间的关系的看法，与作者十几年前在《数学与经济》一书中所表达的看法相比，已经有一些发展。妨碍经济学应用数学方法来研究的并非只有“道德规范”之类的意识形态问题，即使是纯粹的数量分析问题，至少在目前，还有逻辑上的困难。但是对于数学公理化方法作为“理论的标志”，作者始终持肯定态度。某些高深的数学工具使不少经济学工作者望而生畏当然不能作为“数学不该用”的理由。说到底，实践是检验真理的唯一标准。Black–Scholes–Merton 理论载入史册最终还是由于衍生证券市场发展的结果。如果某位经济学家说，因为他不懂随机微分方程，所以这种理论就不能是经济学，那实在是太可笑了。

作者的某些看法受到科普畅销书 Paulos [18] 的影响，也受到张圣平 [7] 的启发。[18] 的中译本被编辑定名为《跨越缺口》，含义是人们应该努力跨越“文学文化”与“科学文化”之间的缺口。正是在这样的意图下，作者希望摆脱常规金融经济学教材 (如 [10]–[11]，[13]–[16]) 的扳起一丝不苟的数学面孔的写法。不过这些常规著作中有时也会露出一点笑脸，我们的 1.6 节中的简单模型其实来自 Duffie [10] 的写法。

Ingersoll [15] 的 81 页上提到了“公共知识”的概念。这是值得注意的。但是其他金融经济学教科书中都没有提到这点。附录中的关于“公共知识”的寓言则并非是 Paulos 的创作。这个寓言在博弈论界流传已久，曾出现在一些博弈论的标准著作 (例如 [17]) 中。但通常说的都是似乎更符合现实的“大男子主义”。

思考题和习题

1. 金融经济学理论是怎样发展起来的？
2. 什么是数学公理化方法？你对它有什么看法？
3. 请把最后 1.6 和 1.7 节中的例子都加入无风险利率 R, 即“当前”的一元钱 (存在银行里) 到“未来”就变成了 $(1+R)$ 元，那么四个例子的计算结果将是怎样的？
4. 完整的无套利假设能否为股票定价？尤其是在所提到的例子中，如果我们只知道股票的“未来”价格，那么根据无套利假设，你能对它的“当前”价格说些什么？

5. 金融产品的定价问题与“公平赌博”的“定价”问题似乎都是在不确定环境下的定价问题。从所讨论的例子中来看，你认为它们间的主要区别在哪里？

参考文献

[1] 艾克纳，A. S.，1990，经济学为什么还不是一门科学，苏通等译，北京大学出版社。

[2] 布劳格，M.，1990，经济学方法论，黎明星等译，北京大学出版社。

[3] 布劳格，M. 等著，R. E. 巴杰豪斯编，2000，经济学方法论的新趋势，张大宝等译，经济科学出版社，北京。

[4] 博兰， L. A., 等著，2000，批判的经济学方法论，王铁生等译，经济科学出版社，北京。

[5] 波珀，K. R.，1999，科学发现的逻辑，查汝强、丘仁宗译，北京经济学院出版社。

[6] 凯恩斯，J. M.，1988，就业、利息和货币通论，徐毓(木丹)译，商务印书馆。

[7] 张圣平，2002，偏好、信念、信息与证券价格，上海人民出版社，上海。

[8] Cochrane, J. H., 2001, Asset Pricing, Princeton University Press, Princeton, N. J..

[9] Dixit, A. K. & R. S. Pindyck, 1993, Investment under Uncertainty, Princeton University Press, Princeton, N.J..

[10] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[11] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N.J..

[12] Friedman, M., 1953，The methodolog of positive economics, in *Essays in Positive Economics*, (University of Chicago Press, Chicago), 3–43 (中译文见：弗里得曼文萃，2001，胡雪峰、武玉宁译，首都经济贸易大学出版社，北京，118–165)。

[13] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[14] Hull, J. C., 1999, Options, Futures, and Other Derivatives, 4th ed., Prentice–Hall, N.J.. (第三版有中译本：赫尔，J., 2000，期权、期货和衍生证券，张陶伟译，华夏出版社，北京).

[15] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[16] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[17] Myerson, R. B., 1991, Game Theory: Analysis of Conflict, Harvard University Press, Cambridge.

[18] Paulos, J. A., 1998, Once upon a Number, Basic Books. L.A.. (中译本：保罗斯，J. A., 2001，跨越缺口，史树中等译，上海科技出版社。)

[19] Pliska, S. R., Introduction to Mathematical Finance: Discrete Time Models, Blackwell, Massachusetts.

附录 (摘自 [18])：狂怒的大女子主义者的寓言和股票市场

我写这个寓言是在 1997 年 10 月股市大跌的一个星期之后。它发生在一个地点不明的愚昧的大女子主义村子里。在这个村子里，有 50 对夫妇，每个女人在别人的丈夫对妻子不忠实时会立即知道，但从来不知道自己的丈夫是否忠实。该村严格的大女子主义章程要求，如果一个女人能够证明她的丈夫不忠实，她必须在当天杀死他。又假定女人们都赞同这一章程，并且都很聪明，也都能意识到别的女人的聪明；同时，还都很仁慈，即她们从不向那些丈夫不忠实的女人通风报信。假定在这个村子里发生了这样的事：所有这 50 个男人都不忠实，但没有哪一个女人能够证明她的丈夫的不忠实，以至这个村子能够快活而又小心翼翼地一如既往。有一天早晨，森林的远处有一位德高望重的女族长来拜访。她的诚实众所周知，她的话就像法律。她暗中警告说村子里至少有一个风流的丈夫。这个事实，根据她们已经知道的，只该有微不足道的后果，但是一旦这个事实成为公共知识，会发生什么？

答案是，在女族长的警告之后，将先有 49 个平静的日子，然后，到第 50 天，在一场大流血中，所有的女人都杀死了她们的丈夫。要弄明白这一切是如何发生的，我们首先假定这里只有一个不忠实的丈夫

A 先生。除了 A 太太外，所有人都知道 A 先生的背叛，因而当女族长发表她的声明的时候，只有 A 太太从中得知一点新消息。作为一个聪明人，她意识到如果任何其他的丈夫不忠实，她将会知道。因此，她推断出 A 先生就是那个风流鬼，于是在当天就杀了他。

现在假定有两个不忠实的男人，A 先生和 B 先生。除了 A 太太和 B 太太以外，所有人都知道这两起背叛，而 A 太太只知道 B 太太家的，B 太太只知道 A 太太家的。A 太太因而从女族长的声明中一无所获。但是第一天过后， B 太太并没有杀死 B 先生，她推断出 A 先生一定也有罪。B 太太也是这样，她从 A 太太第一天没有杀死 A 先生这一事实得知，B 先生也有罪。于是在第二天，A 太太和 B 太太都杀死了她们的丈夫。

如果情形改为恰好有三个有罪的丈夫，A 先生、B 先生和 C 先生，那么女族长的声明在第一天不会造成任何影响，但类似于前面描述的推理过程，A 太太、B 太太和 C 太太会从头两天里未发生任何事推断出，她们的丈夫都是有罪的，因而在第三天杀死了他们。借助一个数学归纳法的过程，我们能够得出结论：如果所有 50 个丈夫都是不忠实的，他们的聪明的妻子们终究能在第 50 天证明这一点，使那一天成为正义的大流血日。

现在我们把森林远处来的女族长的警告代替为对去年 (1997) 夏天泰国、马来西亚和其他亚洲国家的通货问题的警告；妻子们的紧张和不安代替为投资者的紧张和不安；妻子们只要自己的“公牛”没有被刺伤就心满意足代替为投资者们只要自己的“公牛”没有被刺伤就心满意足；杀丈夫代替为抛股票；警告和杀戮之间的 50 天间隔代替为东亚问题和大崩盘之间的延迟，你就会得到这次大崩盘的成因。更清楚地说，利益息息相关的金融集团们可能已经在怀疑其他的亚洲经济是不堪一击的，但直到某人如此公开地说，并最终发觉了他们自身的不堪一击以前，他们是不会行动的。这样，马来西亚总理在 1997 年 4 月批评西方银行的讲话就起着女族长的警告那样的作用，促成了他最担心的这次危机。

幸好不像是故事中的丈夫们那样，市场是能够再生的。华尔街波涛后来的此起彼伏说明，如果妻子们能够让丈夫们在炼狱中短暂停留之后再复活的话，这种类比就会更加逼真。这就是地球村中的生与死、买和卖。

二、二期证券市场的基本模型和线性定价法则

本讲要求：为二期证券市场给出一个基本概率论数学模型，并由此讨论无套利假设的各个层次。线性定价法则将作为无套利假设的一个层次给出。在一定的假设下，指出线性定价法则意味着随机折现因子的存在。而由随机折现因子可导出线性定价法则与资本资产定价模型 (CAPM) 是等价的。

数学预备知识：初等概率论。线性代数。

内容：

- 2.1 无不确定性的无套利假设定价法则。
- 2.2 带不确定性的无套利假设定价法则。
- 2.3 二期证券市场的基本模型。线性定价法则和随机折现因子。
- 2.4 随机折现因子的初步讨论。无风险证券及其模仿组合。
- 2.5 收益超平面和超额收益子空间。
- 2.6 资本资产定价模型与线性定价法则的等价性。
- 附录：数学预备知识 1。

在上一讲中，我们已经指出，金融经济学的基本问题是在不确定市场环境下对金融资产定价。这大致可表达为这样的一个问题：已经知道一种金融资产在未来各种可能的价值，要问它当前的价值是多少。所谓金融资产是相对于实物资产而言的，它们是一些与实物关系不大的合约 (证券)，但是都能用钱来结算。金融资产交易都是为了以钱生钱。但是投入的钱是当前确定的钱，产出的钱则是未来的不确定的钱。因此，人们面临的基本决策问题经常就是未来的不确定的钱在当前究竟值多少钱。例如，上一讲中说到的欧式期权，在它的到期日，它的价格取决于那时的不确定的股票价格。期权定价问题就要问怎样根据它的不确定的到期日价格来定出它的当前价格。又如，各种债券在它的到期日的价格虽然是预先规定的，但是由于在它未到期时市场利率会有随机变化，它的到期日面值的实际价值其实也是不确定的。它的当前定价问题仍然是同类型的。这种问题与普通实物资产定价问题的根本不同在于：对于实物商品来说，由于投入与产出是不同的商品，有时还可通过生产成本加一定的利润来估计其价值，而对金融商品则不可能这样做。

对金融资产定价的一种朴素的想法是：估计其未来价值的各种可能性，然后计算其平均值，并再以这一平均值来作为它的估值。或者用数学语言来说，把金融资产的未来价值看作一个随机变量，那么这

个随机变量的数学期望就可当作该金融资产的当前价值。具体应用时，只要通过一定的方法来估计未来价值的概率分布，然后算出其数学期望，就认为问题已经得到解决。这样的想法如同概率论这门学科本身一样古老。众所周知，概率论起源于对一些赌博问题的研究，而对赌博问题的定价就是这样解决的。在金融经济学的早期研究中，人们也会很自然地这样来认识金融资产定价问题。而金融经济学发展中走出的最重要的一步是否定了这样的想法，而提出用无套利假设来定价；在期权定价问题中，完整的无套利假设将为未来价值这一随机变量赋予特殊的 (不同于客观可能性的) 概率分布，它们之间的关系就称为资产定价基本定理。

在上一讲中，我们也已经用粗浅的数学模型来描述怎样利用数学来解决这样的问题。但是在那里，我们为了便于读者接受，没有引进比较精确的数学语言。对于反映“无钱投入就无钱产出”的无套利假设在没有明确的数学框架的情况下，表达上也有很多模糊的地方。在这一讲中，我们就要更讲究数学形式化，使其中的表达不再有任何含糊。不过在这一讲的讨论中，我们暂且把要求更高的完整的无套利假设放在一边，而先讨论一种层次稍低的无套利假设。它就是“线性定价法则”。这一定价法则的作用很大，它实际上可用来概括 Black–Scholes 期权定价理论以前的金融经济学的几乎所有主要结果。

2.1　在我们当前的讨论中，无套利假设的一些内容将作为“公理”来引入。我们的首要问题是怎样把它数学模型化。为了便于理解，我们先讨论没有不确定性的模型，再讨论有不确定性的模型。讨论没有不确定性的模型，像是讨论一个没有随机变化的债券市场，或者说像讨论一个储蓄问题，其结果是一目了然的。然而，即使如此，也已经能给人许多启发。

在我们紧接的一些叙述中，我们都只讨论二期模型，即从时间的角度来看，假定我们的模型中只有“当前”与“未来”两个**时刻**(为方便起见，以后我们把引号省略)。所谓**证券市场**是由 K 种 (**基本**) **证券**及其各种组合所组成。这些证券被假定为未来价格已知的一种商品。所谓证券定价问题，就是要根据证券未来的价格以及一些假设来确定它当前的价格。这 K 种证券可以用代表它们的未来价格的正实数 $x_1, \ldots, x_K > 0$ 来表示。其总体则可看作一个 K 维列向量 $x = (x_1, \ldots, x_K)^T$, 这里 T 表示转置。这样做自然已经包含许多经济学上的假设：价格可以用正实数来表示，而不是如日常生活中那样，任何价格都有一个最小货币计量单位；市场上只有一种大家都接受的价格，而不是如日常交易中往往存在许多种价格；价格是正实数，而不考虑有非正价格的证券 (一钱不值的证券其实也是司空见惯的)；如此等等。不过以下的讨论对价格是否为正并无特别要求。尤其是一个证券组合的价值就不一定是正的。

而在以后的理论框架中，证券与证券组合已经没有本质区别。

现在来考虑证券投资。证券投资行为可由一个 K 维行向量 $\theta = (\theta_1, \ldots, \theta_K)$ 来表示，这里 θ_1 表示买入第 1 种证券的数量，θ_2 表示买入第 2 种证券的数量，如此等等，以至 θ_K 表示买入第 K 种证券的数量。这里为了数学上处理方便起见，假设这些 θ_i 都可以是任意实数，而不是像实际生活中，证券的数量都只能是整数；同时，当 θ_i 是负数时，这里的“买入”其实是“卖出”。尤其是 θ 的相反向量 $-\theta = (-\theta_1, \ldots, -\theta_K)$ 是表示与 θ 正好相反的投资活动。这样的 K 维向量经常称为（**投资**）**组合** (portfolio) 或者（**投资**）**策略** (strategy)。因为每种证券在未来的价格是已知的，所以一个投资组合或者一项投资策略 θ 的未来价值就被合理地假定 (这其实也是一种“线性定价法则”) 为行向量 θ 与列向量 x 的乘积：

$$\theta \cdot x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_K x_K. \tag{1}$$

由于 θ_i 可以取负值，这个数值也是可正可负的实数。这样，所谓证券市场就是组合或策略 θ 的全体。如果市场上对投资策略没有任何限制，那么 θ 可取任何 K 维向量。于是所谓证券市场，用数学的语言来说，也就是一个实数域上的 K 维向量空间。我们把它记作 M.

至于**定价问题**是这样来提的：已知一个投资组合未来的价值，要确定该组合当前的价值。由于在我们目前的证券市场模型中没有任何经济学内容 (例如没有投资－消费者，没有生产，没有消费，没有交易等等)。因此，我们不可能在这里通过各种宏观经济关系来为证券定价。我们能做的只是“相对定价”，即如何根据基本证券的价格，来为证券组合定价。

现在我们来把反映“无钱投入就无钱产出”的模糊的无套利假设精确化。我们的问题框架已经变为一个 K 维向量空间。而定价问题无非是要对这个 K 维向量空间中的每一个向量给出一个实数来作为它的当前价值。无套利假设则是要对这种定价定出若干法则。这样的法则按其要求的不断提高，可分为这样五个层次：

1. 未来价值一样的组合，当前应该有一样的定价。这是因为如果有两个组合 θ 和 ξ 的未来价值 (请注意，我们这里有一个先决条件是组合的未来价值是由 (1) 定义的。这个定义也是一个重要假设) 是一样的，而它们当前的价值则是 θ 比 ξ 要便宜，那么投资者就可通过买进一个 θ, 卖出一个 ξ 在当前赚一笔钱，而当未来交割时他并无交割的困难 (这里自然又对交易作出一些假设，即当前“卖空”一个组合可到未来再交割)。这就使他获得了套利。从而违背了无套利假设。由于对应每个组合的未来价值，只对应其唯一的当前价值，定价问题就变为如何来确定一个**定价函数**，它是一个实变量的实值函数：$p{:}\,\mathbb{R} \to \mathbb{R}$, 其中 $\mathbb{R}$ 表示实数全体。p 的自变量是组合的未来价值，因变量是组合的当

前价值。请注意，p 的自变量不是 M 中的元素，即 p 不是证券组合 (K 维向量) 的函数，而是证券组合未来价值 (实数) 的函数。

2. 组合的若干倍的当前价值应该等于该组合的当前价值的同样倍数。这是因为如果这两个量不相等的话，如同组合的“批发价”与“零售价”不同。于是人们就可利用其中的差价，运用比如“批发”买入，“零售”卖出那样的策略，获得套利。

3. 组合的买价与卖价应该一致。在我们讨论的市场模型中，不涉及交易所、做市商之类的市场体制。因此，市场上的证券组合的买价与卖价应该一致。否则人们就可通过买卖差价来获得套利。

4. 组合的当前价值应该等于其组合成分的当前价值之和。这一规定的理由与第 2 点类似。否则人们可以利用其中的差价，运用比如“一揽子”买入组合，“分散”卖出组合成分的策略，来获得套利。

5. 未来值钱 (价值为正) 的组合，当前也值钱。这一规定如果不成立，就会出现当前不花钱买入一个组合，未来白赚一笔钱。与前面不同的是，前面三个层次的假设不成立，所获得的套利是在同一个时期得到的。而这一层次的假设不成立，套利可能是跨时期获得的。不妨称后一种套利是“跨时套利”。

这五个不同层次的无套利假设虽然看起来已经很清楚，但是如果用数学形式来表达，那就更为简明确切。

无不确定性的无套利假设定价的五个层次：

1. **(可定价法则)** *存在定价函数* $p\colon \mathbb{R} \to \mathbb{R}$.
2. **(正齐次定价法则)** p *是正齐次函数，即对于任何正实数* $\lambda > 0$ *和实数* y, *有* $p(\lambda y) = \lambda p(y)$.
3. **(齐次定价法则)** p *是齐次函数，即对于任何实数* λ *和实数* y, *有* $p(\lambda y) = \lambda p(y)$.
4. **(线性定价法则)** p *是线性函数，即对于任何实数* λ, μ *和任何实数* y, z, *有* $p(\lambda y + \mu z) = \lambda p(y) + \mu p(z)$. *这样的定价函数一定有这样的形式：*$p(y) = ay$, *其中* a *是实数。*
5. **(正线性定价法则)** p *是正线性函数，即* p *是线性函数，并且当* $y > 0$ *时，*$p(y) > 0$. *这样的定价函数一定有这样的形式：*$p(y) = ay$, *其中* $a > 0$.

在一般的证券市场理论框架下，这五个层次的无套利假设似乎都应该满足。但是我们也很容易找到不满足这些假设的市场现实。例如，

由于考虑交易费用之类的情况，同一种证券可能有多种价格。这就不满足这里的“可定价法则”。“批发价”与“零售价”不同、买卖价不同 ($p(y) \neq -p(-y)$) 或“一揽子价”与“分散价”不同，就不满足“正齐次定价法则”、“齐次定价法则”或“线性定价法则”。存在“跨时套利”时，就不满足“正线性定价法则”。如此等等。以前在理论上讨论金融问题时，把市场理想化，总认为包含五个层次的完整的无套利假设都应该满足。但是现在研究越来越深入，我们就可把层次分得更加细些。正如我们在上一讲中指出，Markowitz 证券组合选择理论、资本资产定价模型 (CAPM) 和 Ross 的 APT 理论其实都只需要线性定价法则，甚至我们以后可以证明，在很一般的未来价值不确定的情形下，前两者当假定正齐次定价法则成立时都与线性定价法则等价。这就意味着这些理论并不要求“跨时套利”不存在，从而也适用于存在“跨时套利”的某些行为金融学之类的研究。这些都是后话。我们暂时还是先回到我们的简单模型上来，那就是证券组合未来价值是确定实数的情形。

现在我们先假定 (可定价法则和) 正齐次定价法则成立。在这一条件下，我们可以定义一个证券组合的收益率。设 $\theta = (\theta_1, \ldots, \theta_K)$ 为市场 M 中的一个组合，其未来价值 $p(\theta \cdot x) \neq 0$. 那么 $r_\theta = \theta \cdot x / p(\theta \cdot x)$ 就称为 θ 的**收益率**。特别是，证券 x_i (前面已经提到，我们以后总用第 i 种证券的未来价值来表示该证券自身) 的收益率为 $r_i = x_i / p(x_i)$. 其含义是组合的未来价值与当前价值之比。请注意，这里所说的收益率与习惯上使用的收益率有所不同。习惯上使用的收益率经常是组合的未来价值和当前价值之差与当前价值之比，它与上述定义相差 -1. 但是在理论讨论上，使用我们目前的定义有许多方便之处。因此，我们以后就采用目前的定义。读者在应用本讲义 (以及许多采用同样定义的著作) 中有关收益率的结果时，必须注意到其中的差别。还有要注意的是，*由于我们假定正齐次定价法则成立，一个证券组合的收益率与它的倍数组合的收益率是一样的*。这一点是符合“常识”的：两个投资者购买同样的证券，那么他们的投资收益率是一样的，而与他们的投资量无关。

当前价值为正的组合的收益率本身也可看作某个组合的未来价值。其特征是它所对应的当前价值一定为 1. 这是因为由正齐次定价法则，

$$p(r_\theta) = p\left(\frac{\theta \cdot x}{p(\theta \cdot x)}\right) = \frac{p(\theta \cdot x)}{p(\theta \cdot x)} = 1.$$

这样一来，我们可以定义 *(无风险) 收益率是当前价值为 1 的未来价值*。这一定义与我们平时理解的收益率似乎有点距离，其实正是由于如上定义收益率，数学上才有这样的方便。如果按习惯来定义收益率，它就没有这样的好处。即使线性定价法则成立，也会涉及一项处理起来不太方便的 $p(1)$. 以后，我们就采用这个收益率定义。

下面的简单定理是一目了然的，但是它具有重要意义。

定理 2.1 在线性定价法则假设下，任何 (非零无风险) 证券组合的收益率都相等。

证明 设定价函数 p 满足 $r=1/p(1)$. 第 i 种证券的收益率为 r_i. 那么

$$r_i=\frac{x_i}{p(x_i)}=\frac{x_i}{x_ip(1)}=1/p(1)=r.$$

即 r_i 与 r 相等。对于一般的证券组合 $\theta=(\theta_1,\ldots,\theta_K)$, 其收益率 r_θ 为

$$r_\theta=\frac{\theta_1x_1+\cdots+\theta_Kx_K}{p(\theta_1x_1+\cdots+\theta_Kx_K)}$$

$$=\frac{\theta_1x_1+\cdots+\theta_Kx_K}{(\theta_1x_1+\cdots+\theta_Kx_K)p(1)}=r,$$

其中要求 θ 的未来价值不为零。 □

这一定理的推理虽然非常简单，但其结论却非常重要。常值 $r=1/p(1)$ 称为**利率**。不过我们上面的结论是在一个完全没有不确定性的模型中得到的。如果模型中有不确定性，即存在某些其未来价格有多种可能性的证券，那么我们把其中的无不确定性 (即“无风险”) 的证券分离开来构成一个子模型，仍能由线性定价法则得到类似的结论：任何无风险证券组合的收益率是一样的。这时相应的 $r=1/p(1)$ 将更确切地称为**无风险利率**。在上一讲中我们讨论期权定价公式时，第一种观点的根据就在于此（那里的无风险利率假定为 1)。历史上，这一结论确实也是 Black–Scholes 期权定价公式证明中的关键。

2.2 现在我们来讨论一个未来带不确定性的证券市场模型。仍假设模型中有 K 种证券。这些证券的未来价格可能都是不确定的。证券的未来不确定价格用概率论中的**随机变量**来描述。这里当然也意味着某种假定，因为除了用随机变量来描述不确定性以外，我们也可能用其他数学概念来描述不确定性。但是概率论是发展最成熟的数学学科。迄今为止，金融经济学研究中的不确定性，还很少不用随机变量来刻划。我们这里所说的随机变量就是初等概率论中定义的随机变量，它可能取任何实数值，而它取某个实数区间的值的概率可用该随机变量的分布函数来刻划。在一些更简单的模型中，常假定未来只能有有限种状态。例如，在上一讲的简单模型中，未来只有两种可能。这时所涉及的随机变量也只可能取两个不同的值。在这种情况下，数学上会简单很多。但在目前的讨论中我们不需要这一简化假定。

与以前一样，这 K 种证券的未来价格仍表示为 $x_1,\ldots,x_K \gg 0$, 但这 K 个未来价格都是 (正) 随机变量，而一般不是普通正实数。记号 $x \gg 0$ 表示 x 取正数的概率为1. 不过以后我们很少用它。记号 $x > 0$ 则留给大于零的概率为正的非负随机变量。如果 x_i 是正实数，或者更确切地说，是定常正随机变量，那么这就意味着其未来价格是确定的。因此，该 x_i 就代表一种无风险证券。然而，在我们下面的讨论中，不假定市场中存在无风险证券，或者存在无风险证券组合。这样的讨论就更具有一般性。K 种证券的总体用它们的未来随机价值来表示，则可看作一个 K 维随机列向量 $x = (x_1,\ldots,x_K)^T$. 与以前的不同在于，这个向量的分量都是随机变量。至于证券组合或投资策略，也仍然是一个 K 维行向量 $\theta = (\theta_1,\ldots,\theta_K)$, 其意义也与以前一样。其分量可以是确定的量，也可以是随机变量。这要看讨论的问题是什么来决定。不过我们在这里定义证券市场 M 时，只考虑确定性的组合，即只考虑 θ_i 是普通实数的情形。这样的 θ 的全体仍然是一个实数域上的 K 维向量空间。它与上节中定义的 M 本质上是一样的。但一个组合 θ 的未来价值将是定常行向量 θ 与随机列向量 x 的乘积：

$$\theta \cdot x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_K x_K. \tag{2}$$

这个数值则是一个随机变量，它反映组合的未来价值的不确定性。市场 M 中的组合的未来随机价值所形成的随机变量全体 $\mathcal{M}$, 不妨称它们为**可交易的未定权益**，即 $\mathcal{M}$ 定义为

$$\mathcal{M} = \{y \,|\, \exists\theta \in M,\ y = \theta \cdot x.\}$$

“未定权益 (contingent claim)”这个词意味着其未来价值是不确定的。而“可交易的”是指这一未定权益可以与市场 M 中的某个组合相对应。在所研究的问题的一定框架下，如果所涉及的未定权益都是可交易的，这种市场就是所谓“完全市场”。这时，如果再有线性定价法则，那么每一未定权益就都可通过基本证券的定价来定价。这种定价的思路正是我们在上一讲中提到过的第二种观点。

现在我们来明确目前情形下的无套利假设定价法则。这里基本上可原封不动地照搬。唯一不同的是目前的定价函数将不再是代表确定的组合未来价值的实数的函数，而是“可交易的未定权益”的未来随机价值全体 $\mathcal{M}$ 上的函数。具体地说，它可以表达为

带不确定性的无套利假设定价的五个层次：

1. **(可定价法则)** *存在定价函数* $p\colon \mathcal{M} \to \mathbb{R}$.
2. **(正齐次定价法则)** p *是正齐次函数，即对于任何正实数* $\lambda > 0$, *有* $p(\lambda y) = \lambda p(y)$.

3. **(齐次定价法则)** p 是齐次函数，即对于任何实数 λ, 有 $p(\lambda y) = \lambda p(y)$.
4. **(线性定价法则)** p 是线性函数，即对于任何实数 λ, μ 和任何 $y, z \in \mathcal{M}$, 有 $p(\lambda y + \lambda z) = \lambda p(y) + \mu p(z)$.
5. **(正线性定价法则)** p 是正线性函数，即 p 是线性函数，并且当 $x > 0$ 时，$p(x) > 0$. 这里 $x > 0$ 表示 x 为非负随机变量，并且它为正的概率大于零。

虽然这里与前面的区别似乎只有一处，即定价函数 p 的定义域从实数域 $\mathbb{R}$ 变为可交易的未定权益全体 $\mathcal{M}$, 但是其数学上的差别还是要注意的。其中最重要的是 $\mathcal{M}$ 有向量空间的结构。事实上，按照它的定义，不难证明，如果 $y, z \in \mathcal{M}$, 那么对于任何 $y, z \in \mathcal{M}$ 和任何实数 λ, μ, 都有 $\lambda y + \lambda z \in \mathcal{M}$. 因此，后面的三条定价法则都有意义。

下面我们就要根据这样的理论框架来讨论金融资产定价问题，或者更确切地说，讨论未定权益的定价问题。我们的讨论只需要线性定价法则。至于正线性定价法则 (它代表“完整的无套利假设”)，在期权定价问题以前，暂时没有用。这里顺便提醒读者，在金融经济学的文献中，“无套利假设”通常就指正线性定价法则。但是也有相当多的文献中，“无套利假设”用来指线性定价法则，其中甚至包括 Modigliani–Miller 和 Merton 的经典文献。因此，在阅读文献时，一定要搞清楚其中的“无套利假设”究竟指什么。正如我们前面提到过的，在 Cochrane [2] 中把线性定价法则称为“单一价格法则”，正线性定价法则称为“无套利”。这可能会流传很广。但是“单一价格法则”这一术语容易引起误解。本讲义将始终采用这里的“五个层次”的说法。

2.3 现在我们将把线性定价法则作为“公理”来建立未定权益定价问题的理论框架。为此，我们还需要把上面已经建立的框架进一步精确化。首先我们可注意到的是，在这些定价法则的定义中，“可交易的未定权益”全体 $\mathcal{M}$ 似乎比证券市场 (证券组合全体) M 更重要，因为在最后的定义中，我们只看到 $\mathcal{M}$, 而 M 已经消失。这样，我们可以直接从随机变量的集合 $\mathcal{M}$ 出发来讨论问题，不必再提及 M. 重要的是 $\mathcal{M}$ 要有向量空间的结构，即如果 $y, z \in \mathcal{M}$, 那么对于任何实数 λ, μ, $\lambda y + \mu z \in \mathcal{M}$.

在这样的思路下，我们说一个证券市场中的未定权益空间，是指随机变量所形成的一个 (实) 向量空间。“基本证券”则是指这个未定权益空间中的一个集合。而“可交易的未定权益”则是指它是某些基本证券的线性组合 (这里的组合应该是 combination, 但是把它理解为 portfolio, 也没有错)。这样建立的框架比前面要广。首先是“基本证券”的个数

不一定是有限种，它也可以有无限种。甚至这无限种“基本证券”相互间还可以是线性无关的，即谁也不是其他证券的组合。在这种情况下，未定权益空间将是一个无限维向量空间。无限维向量空间的概念虽然不在初等线性代数教程中讲授，但是用有限维向量空间的理论去理解它也不会有多少困难。“基本证券”的个数无限在实际市场中是有可能的。例如在证券市场中把时间因素、随机因素都考虑进去，不同时间、不同场合发行的同一公司的股票都可认为是不同的证券，那么认为市场上有无限种证券比认为只有有限种证券要更合理一些。以后我们一般并不假定“基本证券”只有有限种。那样做其实也不会给我们多少数学上的困难。

有了未定权益空间以后，我们就可对未定权益讨论定价法则。如果线性定价法则成立，而“基本证券”又都可以定价，并且符合线性定价法则，那么所有可交易的未定权益也都可以唯一定价。如果可交易的未定权益的全体就等于整个未定权益空间，那么所有未定权益都可以唯一定价。这时，就称该证券市场是**完全市场**，否则就称为**不完全市场**。在这个框架中，不完全市场中的未定权益定价问题就变为怎样把“可交易的未定权益空间”中的定价函数扩充到整个未定权益空间中去。在数学上，它可表达为这样一个问题：一个小向量空间中的线性函数怎样扩充到一个包含它的大向量空间中去。对于这个问题，我们只要设想一下，一个单变量的线性函数怎样扩充为二变量的线性函数，就可理解，作为一个数学问题它有无限多个解，并且这些解可以很随意。这对一个金融问题来说不是一个好答案。因此，线性定价法则对于不完全市场的未定权益问题来说是不够的。这时，正线性定价法则就能给出更合理的解答。但是对于完全市场来说，线性定价法则已经足够。这就是为什么我们可以先局限在线性定价法则的假定下进行许多重要的讨论。以上这些话看来平淡无奇，其数学内容也极为简单。然而，它确实就是金融经济学数学框架中的核心。我们在以后将越来越体会到这一点。

不过话又说回来，上面的框架毕竟又太一般。这样一般的框架中很难得出一些有金融意义的结果。于是我们又需要新的假设。这个新的假设是：*未定权益随机变量的方差有限*。一个随机变量的方差刻划的是该随机变量可能有的变化的幅度。所谓“未定权益随机变量”就是我们在当前看一个证券或证券组合的未来不确定价值。如果我们把它作为一个随机变量来看，那么假定它的方差有限是合理的。否则将意味着未来不确定价值的随机变化毫无节制。在这样假定下，我们可以在未定权益向量空间，引进新的数学结构。即对于任何 $y, z \in \mathcal{M}$, 定义它们的**内积**为 y, z 的二阶矩 $E[yz]$, 这里 $E[\cdot]$ 表示随机变量的数学期望。那么它就形成一个内积空间。我们再进一步的假定是：这个内积空间是完备的，即它是一个 **Hilbert 空间** (有关的数学知识见附录)。

为说明对于方差有限的随机变量这样做是合理的，我们先来回顾一下一个随机变量的协方差和方差的定义。设 y, z 为两个随机变量。则

$$\mathrm{Cov}[y, z] = E[(y - E[y])(z - E[z])] = E[yz] - E[y]E[z]$$

称为 y 和 z 之间的**协方差**。当 $y = z$ 时，

$$\mathrm{Var}[y] = \mathrm{Cov}[y, y] = E[(y - E[y])^2] = E[y^2] - E^2[y]$$

称为 y 的**方差**。对于任意的随机变量来说，它的数学期望和方差都不一定存在。不过那种随机变量都是“病态”的。我们现在假定未定权益的方差有限，是指它的数学期望和方差都存在，并且都取有限值。

两个随机变量的方差和协方差之间有一个重要的 Cauchy 不等式：

$$\mathrm{Cov}^2[y, z] \leqslant \mathrm{Var}[y] \cdot \mathrm{Var}[z],$$

其中等号当且仅当 $y = az + b$ 时成立，这里 a 和 b 为任意常数。因此，当 y 和 z 的方差都有限时，它们之间的协方差也有限，从而再由协方差的定义，就可得到 y 和 z 的二阶矩 $E[yz]$ 也有限。这样在由方差有限的随机变量所构成的向量空间中，就可对其中的任何两个元素 y, z 用 $E[yz]$ 来定义它们的内积。容易验证，这样定义的内积是符合内积的要求的。内积的好处是引入了一种数学结构。更确切地说，内积使向量空间中有了角度的概念，尤其是有了两个向量“正交”的概念。具体地说，对于上面定义的未定权益 Hilbert 空间 $\mathcal{M}$ 来说，所谓 $y, z \in \mathcal{M}$ 正交，是指 $E[yz] = 0$.

“未定权益空间是 Hilbert 空间”这一假设在未定权益空间是由有限个基本证券生成时一定成立，因为这时所形成的向量空间一定是有限维的，而有限维的 Hilbert 空间就是所谓“欧氏空间”，它一定是完备的。但是我们有时也要讨论无限维的未定权益空间。尤其是在讨论 Ross 的 APT 理论时就是如此。最后我们还需要假定，线性定价函数 $p\colon \mathcal{M} \to \mathbb{R}$ 是连续的。这在 $\mathcal{M}$ 是有限维空间时是不成问题的，因为任何有限维空间上的线性函数一定是连续的。但是当 $\mathcal{M}$ 是无限维空间时，这个假定是必要的。历史上在讨论 Ross 的套利定价理论时，它相当于所谓“渐近无套利假设”。我们将在第四讲向大家介绍这一点。

综上所述，作为我们的理论的出发点的“公理”是这样几条：

基本假设：

1. 未定权益空间 $\mathcal{M}$ 是一些方差有限的随机变量形成的向量空间。
2. 如果对于任何 $y, z \in \mathcal{M}$，定义 $E[yz]$ 为它们的内积，那么 $\mathcal{M}$ 是 Hilbert 空间。

3. 定价函数 $p\colon \mathcal{M} \to \mathbb{R}$ 为线性连续函数。

在这样的假设下，我们立即可得下面的基本定理：

定理 2.2 (随机折现因子存在定理) 在上述基本假设下，存在唯一的 $m \in \mathcal{M}$, 使得对于任何 $y \in \mathcal{M}$, 有 $p(y) = E[my]$.

这里的 m 是未定权益空间 $\mathcal{M}$ 中的元素，即它本身也可以看作一种证券组合。它称为 $\mathcal{M}$ 中的**随机折线因子** (Stochastic Discount Factor, SDF)。

作为一条数学定理来看，随机折现因子的存在是很平凡的，因为它是所谓“Riesz 表示定理”的直接推论。即使是 Riesz 表示定理也很好理解。拿这里的情形来看，我们可以这样来简述定理的证明：设 $R_0 \subset \mathcal{M}$ 为所有当前价值为零的未定权益全体 (以后我们将称它为“超额收益子空间”)，即对于任何 $y \in R_0$, 有 $p(y) = 0$. 那么由于定价函数 p 是个连续线性函数，R_0 构成 $\mathcal{M}$ 的一个闭线性子空间。这而闭子空间一定可以表示为与某个向量垂直 (或“正交”，即与该向量的内积为零) 的向量全体。把这个“垂直向量”作适当的长度上的变化，就可以得到唯一的随机折现因子 m.

然而，作为一条金融经济学的定理，这条定理是意味深长的。它说明，在一个合理的金融经济学的资产定价理论的框架中，任何定价法则，只要它是线性定价法则，那么它就一定对应一个随机折现因子。这个框架目前没有任何经济学内容。但是任何经济学讨论得到的定价法则，只要它仍然是 (连续) 线性的，那么它一定有相应的随机折现因子。例如，古典的“赌博理论”认为，一场公平的赌博就是对参赌者来说未来价值的数学期望为零的赌博。因此，把一场赌博的未来价值的数学期望作为赌博的当前价值，那么其相应的随机折现因子就是常数 1 (如果 $1 \in \mathcal{M}$)。又如，Black–Scholes 期权定价理论中，最后的期权定价是期权的未来价值的对某种“等价概率鞅测度”的数学期望。这种“等价概率鞅测度”使股票的当前价值恰好是其 (折现) 未来价值的数学期望。由此也可定出某种对“客观概率测度”的随机折现因子。其他如每个投资者对市场上的各种基本证券都有根据自己掌握的信息以及偏好的定价。这种定价其实也是给出了一个具有该投资者特征的随机折现因子。行为金融学中的许多讨论也不过是给出与经典讨论不同的随机折现因子。如此等等。

这样一来，我们就建立了一个讨论金融资产定价问题的数学框架。在这个框架中，未定权益这个概念与方差有限的随机变量等同起来。而未定权益又可包括证券、证券组合、期权、衍生证券、金融资产等等概念。还有在我们的讲义可能不用，但是在金融实务的著作中经常使

用的“头寸 (position)”也可列在其中。而随机折现因子存在定理告诉我们，甚至“定价法则”也等同于某个未定权益或方差有限的随机变量。但是这里的所谓“等同”、“包括”都是指它们在**外延**上都是一个方差有限的随机变量，而并不是指它们在**内涵**上也是等同的。即使在数学中，向量与随机变量也决不在内涵上同义，而只是在目前这样的框架中，一个随机变量就可看成一个向量空间中的向量，使得有关向量空间的结果都能搬到这种情况下的随机变量上来。

现代数学家认为一个数学概念是由定义它的公理体系来决定的，而与它叫什么名字无关。大数学家 D. Hilbert (1861–1945) (我们已经遇到过以他命名的抽象向量空间，不过那当然是后人这样命名的) 就说过，在他的几何公理体系中，把“点、线、面”叫成“桌子、板凳、啤酒瓶”，并不影响公理体系本身。但是这句话可能只对数学家有效，因为他们只关心他们所研究的一些概念的相互之间的逻辑关系。而对一个普通人来说，这样一改名以后，他首先会想到的是“桌子、板凳、啤酒瓶”与“点、线、面”的内涵绝然不同。现代科学体系的创立者们在很大程度上接受了数学家的这种坏习惯，使得他们在提出新术语上不大考虑普通人会“望文生义”。结果经常发生的是，这些新术语或者令人不知所云，或者使人随意联想[1]。

至于后面这些金融名词在内涵上更是互不相同，但是在目前的框架中，它们全成了外延上相同的随机变量，并且可以对它们进行向量运算。当然，我们也可以在数学上细化使比如证券与期权有所区别。但是只要注意一个概念的外延和内涵上的限制以及所讨论的问题的主题，这种区别并不是一定要引入的。

2.4 有了随机折现因子以后，我们立即可以得到一些资产定价的基本性质。事实上，由协方差的定义，我们有

$$\mathrm{Cov}[y,m]=E[my]-E[m]E[y].$$

因此，由基本定价公式，可得

$$p(y)=E[m]E[y]+\mathrm{Cov}[y,m]. \tag{3}$$

为了说明这个表示式的经济意义，我们先来引入“无风险证券” (或“无风险权益”)。在上面定义未定权益空间 $\mathcal{M}$ 时我们没有专门提及无风险证券。所谓无风险证券是指其未来价值是确定常数 (或者更确切地说，等于常数的概率为 1) 的证券 (或者更一般地说，“证券组合”、“未定权益”等等)。一个市场中不一定有无风险证券。例如，目前的中国股

---

[1] 一个典型的例子是前些年甚为热闹的“新三论”：系统论、协同论、突变论。不少人根本没有弄清楚它们的内涵外延是什么，就“论述”起来。

市中，就很难找到一种无风险证券 (国外通常可采用短期国库券作为无风险证券)。这样，上面定义的未定权益市场 $\mathcal{M}$ 中也不一定要有无风险证券。换句话说，常数 1 可能不是 $\mathcal{M}$ 的元素。

在通常情况下，无风险证券的作用是为资产定价问题建立一个相对计量单位。就像宏观经济的统计中，我们常常要把货币计量单位折算为某一年的“不变价格”来计算，才能使经济的变化可以相互比较。如果无风险证券 $1 \in \mathcal{M}$, 那么它的价格就是 $p(1) = E[m]$. 我们在以后总假定 $E[m] \neq 0$. 否则会形成“当前不值钱，未来肯定值钱”的“套利权益”。这样，无风险证券的收益率应该是 $1/p(1) = 1/E[m] = r_f$. 以后我们就用 r_f 来表示无风险利率，并且要注意到它是一个相对于某个定价函数来说的常数。它与市场中是否存在无风险证券没有关系。基本定价公式 (3) 因而又可写成

$$p(y) = \frac{E[y]}{r_f} + \mathrm{Cov}[y, m]. \tag{4}$$

这个表示式就把一个证券或一个未定权益的当前价值分解为两部分。前一部份 $E[y]/r_f$ 是它的**时间价值**，即它的未来期望价值对无风险利率的折现；后一部分 $\mathrm{Cov}[y, m]$ 则是它的**风险价值**。它是由于未来价值可能有的随机波动所引起的，可以用来解释为什么股票的当前价值会与债券的当前价值有所不同。这一价值是未来价值 y 与随机折现因子 m 的协方差。这一协方差当然也可能为零。由此我们发现一大类很有意思的未定权益，它们就是无风险价值的未定权益。再考虑到收益率的定义，我们还可发现这类无风险价值的未定权益的期望收益率 $E[r_y] = E[y/p(y)] = r_f$. 这一结果比定理 2.1 的意义更加广泛。

但是如果市场中不存在无风险证券，那么我们该怎样来处理无风险利率及其有关关系呢？这个问题光从金融学的角度来考虑，似乎很难解决。然而，如果从我们的数学框架来考虑问题，我们会发现，由此可得到一个很有意义的概念。

事实上，常数 1 虽然不是 $\mathcal{M}$ 的元素，但是由于 $\mathcal{M}$ 中的元素都是方差有限 (它还蕴含数学期望有限) 的随机变量，我们可以对任何 $y \in \mathcal{M}$ 求它的数学期望 $E[y]$. 这个数学期望，在 $1 \in \mathcal{M}$ 时可看作 1 与 y 的内积。而在 $1 \notin \mathcal{M}$ 时，它又可以看作 $\mathcal{M}$ 上的一个连续线性函数。于是我们再由 Riesz 表示定理可知，存在唯一的元素 $1_{\mathcal{M}} \in \mathcal{M}$, 使得对于任何 $y \in \mathcal{M}$, 有 $E[y] = E[1_{\mathcal{M}} y]$.

这个 $1_{\mathcal{M}}$ 称为**无风险证券的“模仿组合** (mimic portfolio)**”**。它的含义是当市场由若干基本证券生成时，这是个模仿无风险证券功能的证券组合。当然，它本身并不一定是无风险证券，除非 $\mathcal{M}$ 中存在无风险证券。当无风险证券 $1 \notin \mathcal{M}$ 时，这个无风险证券的模仿组合 $1_{\mathcal{M}}$ 在许多地

方都可以起无风险证券的作用。但它的收益率不再是 $1/E[m]$, 而是不确定的 $1_{\mathcal{M}}/E[m]$. 值得注意的是，$1_{\mathcal{M}}$ 有如下性质：$0 < E[1_{\mathcal{M}}] < 1$. 这是因为 $1_{\mathcal{M}}$ 是个非常数随机变量，所以它的方差 $\mathrm{Var}[1_{\mathcal{M}}] = E[1_{\mathcal{M}}^2] - E^2[1_{\mathcal{M}}] > 0$. 但 $E[1_{\mathcal{M}}^2] = E[1_{\mathcal{M}}]$. 因此，上式仅当 $0 < E[1_{\mathcal{M}}] < 1$ 时才有可能。尤其是无风险证券模仿组合的期望收益率要严格小于无风险利率。这是非常有意思的事。它至少在理论上告诉我们，在一个没有无风险证券的市场上，不但存在一个无风险利率，并且还有一个期望收益率不等于无风险利率的无风险证券的替代物。对于中国股市来说，这似乎是一个值得做实证分析的研究课题。

有了无风险证券模仿组合以后，我们还可注意到，未定权益的风险价值也对应某个“模仿组合”。事实上，

$$\mathrm{Cov}[y, m] = E[my] - E[m]E[y] = E[(m - E[m]1_{\mathcal{M}})y].$$

即未定权益的风险价值也是一个未定权益空间上的连续线性函数，其对应的“模仿组合”是 $C_m = m - E[m]1_{\mathcal{M}}$. 因此，所谓无风险价值的未定权益，也就是与 C_m 正交的未定权益。注意到这一点，我们就会发现这种未定权益并不罕见。在“古典赌博论”的情形下，认为未定权益的当前价值就是它的数学期望。这时，$p(y) = E[y]$, $m = 1$ (或者更一般的情形，$m = 1_{\mathcal{M}}$, 但 $E[m] = 1$)。于是任何未定权益都是无风险价值的。在 m 是无风险证券 (m 是常数，但不一定为 1) 或者 m 与 $1_{\mathcal{M}}$ 共线 (即 $m = \alpha 1_{\mathcal{M}}$, α 为实数) 的情形下，仍然如此。在 m 是与 $1_{\mathcal{M}}$ 不共线的风险未定权益的情形下，我们可以注意到，如果 $E[m] \neq 0$, 那么一定存在 $m' \in \mathcal{M}$, 使得 $\mathrm{Cov}[m, m'] = E[C_m m'] = 0$, $E[mm'] \neq 0$, $E[m'] \neq 0$. 事实上，这时 m 与 $1_{\mathcal{M}}$ 的线性组合能张成一个二维空间。$C_m = m - E[m]1_{\mathcal{M}}$ 是这个二维空间中的一个与 m 不共线的向量。因此，在这个二维空间中与 C_m 正交的非零向量 m' (这样的向量当然存在) 一定与 m 不正交 (请读者用严格的线性代数办法证明这点)。即存在 $m' \in \mathcal{M}$, 满足 $E[C_m m'] = E[mm'] - E[m]E[m'] = 0$, $E[mm'] \neq 0$. 这样的向量当然还要满足 $E[m'] \neq 0$.

上面的最后一段论证相当重要。它将是我们在下面证明“零-β 资本资产定价模型 (zero-beta CAPM)”时的关键。下面我们还会给出这段论证的更严格的叙述。这是 Black [1] 在 1972 年证明这条定理的一个关键命题。直到现在为止，我们仍然可以在一些教科书上看到类似命题的复杂证明。而我们用更高的数学观点来看这一问题，就发现这个结果仅仅是一个简单的“平面几何”问题。

2.5　说起收益率，我们再在目前这个数学框架中把它进一步明确。设 $\mathcal{M}$ 为未定权益空间，m 为其随机折现因子。那么对于任何当前价值非零的未定权益 $y \in \mathcal{M}$ ($p(y) = E[my] \neq 0$), 定义其**收益率**为 $r_y =$

$y/E[my]$. 由于 $\mathcal{M}$ 是 Hilbert 空间，故我们也有 $r_y \in \mathcal{M}$, 即 r_y 也是一个未定权益，其特点在于 $p(r_y) = p(y)/p(y) = E[my]/E[my] = 1$, 即任何收益率 r_y 的当前价值总为 1. 反之，任何当前价值为 1 的未定权益，也一定是某个未定权益 (例如自身) 的收益率。这样，我们定义

$$R_1 = \{r \in \mathcal{M} \,|\, p(r) = E[mr] = 1.\},$$

并称它为**收益 (率) 超平面**。超平面是三维空间中的平面概念的推广。三维空间中的平面是二维子空间 (它通过坐标原点) 的一个平移。超平面则是比全空间的维数少一维的 (闭) 子空间的平移。在这里，相应的子空间是

$$R_0 = \{r \in \mathcal{M} \,|\, p(r) = E[mr] = 0.\},$$

即当前价值为零的未定权益全体。任取一个收益率 r_y, 我们立即可以看到，对于每个 $r_z \in R_1$, 都有 $r_z - r_y \in R_0$. 反之也一样。这就是说，$R_1 = R_0 + r_y$, 即它是 R_0 的平移。由于有这样的关系，我们也称 R_0 是**超额收益 (率) 子空间**，因为其中的元素都是相对于某个收益率的差额形成的。这里和以后，我们将经常把“率”字省略。

需要注意的是，收益超平面 R_1 不是线性空间，即如果 $y, z \in R_1$, 我们得不到 $\lambda y + \mu z \in R_1$. 但是如果 $\lambda + \mu = 1$, 那么我们就有 $\lambda y + \mu z \in R_1$. 向量空间中有这样性质的子集称为**仿射集**。超平面是向量空间中“最大的”仿射真子集，即真包含它的仿射集一定是全空间。

收益超平面 R_1 对金融经济学理论特别重要。事实上，早期的金融学研究都不从证券或证券组合的价格出发来研究的，而都从证券或证券组合的收益率出发来研究。Markowitz 证券组合选择理论、资本资产定价模型等都是如此。

2.6 在这一讲的最后一节，我们来指出线性定价法则与资本资产定价模型的等价性。为此，我们先来说说什么是资本资产定价模型。与定价法则不同的是，资本资产定价模型是用证券组合的期望收益率来表示的一种定价公式。这个公式可以这样通过收益超平面 R_1 来表达：存在 $r_u \in R_1$ 和实数 $\alpha \neq 0$, 使得对于任何 $r \in R_1$, 有

$$E[r] - \alpha = \frac{\mathrm{Cov}[r, r_u]}{\mathrm{Var}[r_u]}(E[r_u] - \alpha).$$

我们先来解剖一只麻雀，以了解线性定价法则与资本资产定价模型之间的关系。这个麻雀就是最简单的维数为 1 的随机未定权益空间。

假设 $\mathcal{M}$ 的维数为 1, 那么在 $\mathcal{M}$ 中存在一个非零随机变量 z, 使得任何 $x \in \mathcal{M}$ 都可表示为 $x = \lambda_x z$, 其中 $\lambda_x \in \mathbb{R}$ 是实数。这时，$\mathcal{M}$ 中的非

零定价函数 p 如果是线性的，一定可表示为 $p(x)=E[mx]=\lambda_m\lambda_x E[z^2]$, 其中 $m=\lambda_m z$, 且 $\lambda_m\neq 0$. 考虑 $\lambda_x\neq 0$ 的情形。则 x 的收益率 $r_x=x/p(x)=\lambda_x z/\lambda_m\lambda_x E[z^2]=z/\lambda_m E[z^2]$. 而 m 的收益率 $r_m=m/p(m)=\lambda_m z/\lambda_m^2 E[z^2]=z/\lambda_m E[z^2]$. 即 $\mathcal{M}$ 的"收益超平面" R_1 就只有一个元素。由此可推出相应的"资本资产定价模型"成立，即

$$E[r_x]-\alpha=\beta_{r_x r_m}(E[r_m]-\alpha),\quad \beta_{r_x r_m}=\mathrm{Cov}[r_x,r_m]/\mathrm{Var}[r_m]=1, \tag{5}$$

这里 α 可取任意非零值。

反之，设正齐次定价法则成立，即连续定价函数 p 对于任何 $x\in\mathcal{M}$ 和任何正数 $\lambda>0$, 有 $p(\lambda x)=\lambda p(x)$. 但是不一定有 $p(-x)=-p(x)$. 同时，p 还满足 $p(\pm 1_{\mathcal{M}})\neq 0$. 这时，如果存在 $r_u\in R_1$ 和 $\alpha\neq 0$, 使得对于任何 $r\in R_1$, 有

$$E[r]-\alpha=\frac{\mathrm{Cov}[r,r_u]}{\mathrm{Var}[r_u]}(E[r_u]-\alpha),$$

那么对于 $r=1_{\mathcal{M}}/p(1_{\mathcal{M}})$ 和 $r'=-1_{\mathcal{M}}/p(-1_{\mathcal{M}})$,有

$$\frac{E[1_{\mathcal{M}}]}{p(1_{\mathcal{M}})}-\alpha=\frac{\mathrm{Cov}[1_{\mathcal{M}},r_u]}{p(1_{\mathcal{M}})\mathrm{Var}[r_u]}(E[r_u]-\alpha),$$

$$\frac{E[-1_{\mathcal{M}}]}{p(-1_{\mathcal{M}})}-\alpha=\frac{\mathrm{Cov}[-1_{\mathcal{M}},r_u]}{p(-1_{\mathcal{M}})\mathrm{Var}[r_u]}(E[r_u]-\alpha).$$

从而，

$$E[1_{\mathcal{M}}]-\alpha p(1_{\mathcal{M}})=\frac{\mathrm{Cov}[1_{\mathcal{M}},r_u]}{\mathrm{Var}[r_u]}(E[r_u]-\alpha),$$

$$-E[1_{\mathcal{M}}]-\alpha p(-1_{\mathcal{M}})=-\frac{\mathrm{Cov}[1_{\mathcal{M}},r_u]}{\mathrm{Var}[r_u]}(E[r_u]-\alpha).$$

由此就导得，$p(1_{\mathcal{M}})=-p(-1_{\mathcal{M}})$. 再加上 p 是连续正齐次函数以及 $\mathcal{M}$ 是 1 维的，立即可得 p 是 $\mathcal{M}$ 上的线性函数。

这样，我们就在 1 维随机未定权益空间上，在正齐次定价法则成立和 $p(\pm 1_{\mathcal{M}})\neq 0$ 的假定下，证明了线性定价法则与资本资产定价模型的等价性。

1 维的情况是平凡的。但是从中我们可以看到一个线性定价函数怎样与一个期望收益之间的等式等价。其中以下一些条件是重要的：1. 在定价函数的线性表示中，随机折现因子 m 非零，并且 $E[m]\neq 0$, 从而可求出其相应的收益率 $r_m=m/E[m]$. 而为导出这样的结果，对定价函数的相应假定可通过"p 是正齐次定价函数，且 $p(\pm 1_{\mathcal{M}})\neq 0$"来刻划。2. 涉及的常数 $\alpha\neq 0$. 这个常数当 $\mathcal{M}$ 为多维时，它对应与 r_u 不相

关的收益率 r_v 的数学期望，即 $\alpha = E[r_v] \neq 0$, $\text{Cov}[r_u, r_v] = 0$. 除了这两点以外，其他推导都可顺水推舟而得。

一般的情形当然不应该对 $\mathcal{M}$ 的维数有任何限制。但是如果我们从向量空间的观点来看，这条等价定理其实只涉及未定权益的数学期望和未定权益的当前价值两种量。未定权益的数学期望可看作未定权益与 $1_{\mathcal{M}}$ 的内积，而未定权益的当前价值是未定权益与随机折现因子的内积。再考虑到内积的几何意义实际上是一个向量到另一个向量的射影的长度，这就是说，不管未定权益空间的维数是多少。最后只涉及“无风险证券的模仿组合” $1_{\mathcal{M}}$ 与“随机折现因子” m 所张成的二维空间，其他未定权益都只是被射影到这个二维空间中来考虑。下面证明的对一般情形的等价性定理本质上只是一条二维空间中的定理。因此，为保证空间不退化为 1 维，我们还需要 m 与 $1_{\mathcal{M}}$ 不共线。这一条件可由 m 的期望收益 $E[r_m] = E[m/p(m)] = E[m]/E[m^2]$ 与 $1_{\mathcal{M}}$ 的期望收益 $E[r_{1_{\mathcal{M}}}] = E[1_{\mathcal{M}}/p(1_{\mathcal{M}})] = E[1_{\mathcal{M}}]/E[m]$ 不相等来导出；即 $E[m^2]E[1_{\mathcal{M}}] \neq E^2[m]$. 这个条件在 $1_{\mathcal{M}}$ 为无风险证券 1 时，意味着 m 不是无风险证券。

定理 2.3 设未定权益空间 $\mathcal{M}$ 为方差有限的随机变量所构成的 Hilbert 空间. $p: \mathcal{M} \to \mathbb{R}$ 为 $\mathcal{M}$ 上的连续正齐次定价函数，即对于任何 $x \in \mathcal{M}$ 和任何 $\lambda > 0$, 有 $p(\lambda x) = \lambda p(x)$. 同时假定 $p(\pm 1_{\mathcal{M}}) \neq 0$, 这里 $1_{\mathcal{M}} \in \mathcal{M}$ 是无风险证券 1 (如果 $1 \in \mathcal{M}$) 或无风险证券的模仿组合 (如果 $1 \notin \mathcal{M}$)。定义

$$R_1 = \{r \in \mathcal{M} \mid p(r) = 1.\}.$$

那么下列两个命题等价:

1) 存在唯一的非零 $m \in \mathcal{M}$, 且 $E[r_m] = E[m]/E[m^2] \neq E[r_{1_{\mathcal{M}}}] = E[1_{\mathcal{M}}]/E[m]$, 使得对于任何 $x \in \mathcal{M}$, 有 $p(x) = E[mx]$;

2) (零 β-资本资产定价模型，zero-β-CAPM) 存在 $r_u \in R_1$, $E[r_u] \neq 0$, 使得对于任何 $r \in R_1$, 有

$$E[r] - E[r_v] = \frac{\text{Cov}[r, r_u]}{\text{Var}[r_u]}(E[r_u] - E[r_v]), \tag{6}$$

其中 $r_v \in R_1$, 满足 $E[r_v] \neq 0$, $\text{Cov}[r_u, r_v] = 0$, $E[r_v] \neq E[r_u]$.

特别是，如果市场中存在无风险证券，即 $1 \in \mathcal{M}$, 那么也有 (资本资产定价模型，CAPM)

$$E[r] - r_f = \frac{\text{Cov}[r, r_u]}{\text{Var}[r_u]}(E[r_u] - r_f), \tag{7}$$

其中 $r_f = 1/p(1) = 1/E[m]$ 为无风险利率。

此外，当 1) 或 2) 成立时，可取 $u = am + b1_{\mathcal{M}}$, 其中 $a \neq 0$ 和 b 为任意实数，并且任何有上述形式的 u 的收益率 $r_u \in R_1$ 都满足 2), 其中尤其是 $r_u = r_m = m/p(m)$ 时， 2) 成立。

证明 假定条件 1) 成立。那么对于任何有 $p(x) \neq 0$ 的证券 x 来说，其期望收益率 $E[r]$ 满足

$$E[r] = \frac{E[x]}{p(x)} = \frac{E[x]}{E[mx]}.$$

令 $m' \in M$ 满足 $\mathrm{Cov}[m, m'] = 0$. 当假定 $p(1_{\mathcal{M}}) = E[m] \neq 0$ 时，这样的 m' 是存在的，并且还可要求 $p(m') = E[mm'] \neq 0$, $E[m'] = 1$. 事实上，令 $m' = cm + d1_{\mathcal{M}}$. 则由条件

$$\begin{aligned} E[m'] &= cE[m] + dE[1_{\mathcal{M}}] = 1 \\ E[mm'] &= cE[m^2] + dE[m] = E[m] \end{aligned}$$

立即可得

$$c = \frac{E[m](E[1_{\mathcal{M}}] - 1)}{E[1_{\mathcal{M}}]E[m^2] - E^2[m]}, \qquad d = \frac{E[m^2] - E^2[m]}{E[1_{\mathcal{M}}]E[m^2] - E^2[m]}.$$

这里分母不为零的条件由 $E[r_m] \neq E[r_{1_{\mathcal{M}}}]$ 得到。

同时，对于 $r_m = m/p(m) = m/E[m^2]$, $r'_m = m'/p(m') = m'/E[mm']$ 以及 $r = x/E[mx] \in R_1$, 我们有

$$\begin{aligned} E[r] - E[r_{m'}] &= \frac{E[x]}{E[mx]} - \frac{E[m']}{E[mm']} \\ &= \left[\frac{E[x]}{E[mx]} - \frac{1}{E[m]}\right] - \left[\frac{E[m']}{E[mm']} - \frac{1}{E[m]}\right] \\ &= \frac{E[x]E[m] - E[mx]}{E[mx]E[m]} - \frac{E[m']E[m] - E[mm']}{E[mm']E[m]} \end{aligned}$$

由协方差的定义以及数学期望运算的线性性，并注意到 $\mathrm{Cov}[m, m'] = 0$, 立即可得

$$\begin{aligned} E[r] - E[r_{m'}] &= -\frac{\mathrm{Cov}[x, m]}{E[mx]E[m]} = -\frac{E[m^2]}{E[m]} \cdot \mathrm{Cov}\left[\frac{x}{E[mx]}, \frac{m}{E[m^2]}\right] \\ &= -\frac{E[m^2]}{E[m]}\mathrm{Cov}[r, r_m], \end{aligned} \tag{8}$$

特别是，当 $x = m$ 时，由方差的定义，我们有

$$E[r_m] - E[r_{m'}] = -\frac{E[m^2]}{E[m]}\mathrm{Var}[r_m] \tag{9}$$

比较 (8) 和 (9)，我们立即得到

$$E[r_x] - E[r_{m'}] = \frac{\mathrm{Cov}[r_m, r_x]}{\mathrm{Var}[r_m]} \left(E[r_m] - E[r_{m'}]\right), \tag{10}$$

即对于 $m = u$ 有条件 2) 成立。

此外，我们还可注意到，$E[r_m] = E[m]/E[m^2] \neq 0, E[r_{m'}] = 1/E[m] \neq 0, E[r_m] - E[r_{m'}] = -\mathrm{Var}[m]/E[m]E[m^2] \neq 0.$

反之，假定条件 2) 成立。那么我们首先设法把用收益率表示的等式 (6) 转换为用定价函数和证券的数学期望的等式。为此，我们注意到当 $E[r_v] \neq 0$ 时，我们由条件 2) 可导得定价函数 p 是齐次函数，即它不但是 $\mathcal{M}$ 上的正齐次函数，并且对于任何 $x \in \mathcal{M}$, 满足 $p(x) = -p(-x)$. 这一推导可完全如同本节开始时对一维空间的证明一样进行。

这样一来，我们不妨假设 $r = x/p(x)$, $r_u = u/p(u)$, $r_v = v/p(v)$, $E[x] = E[u] = E[v] = 1$. 否则，我们只需用 $\tilde{x} = x/E[x]$, $\tilde{u} = u/E[u]$ 和 $\tilde{v} = v/E[v]$ 来取代 x, u 和 v, 进行同样的讨论。于是等式 (6) 就变为

$$\frac{1}{p(x)} - \frac{1}{p(v)} = \frac{\frac{E[xu]-1}{p(x)p(u)}}{\frac{E[u^2]-1}{p^2(u)}} \left(\frac{1}{p(u)} - \frac{1}{p(v)}\right).$$

因此，

$$p(v) - p(x) = \frac{E[ux]-1}{E[u^2]-1}(p(v) - p(u)).$$

我们要选取的 $m = a'u + b'1_{\mathcal{M}}$ 是希望它满足 $p(x) = E[mx]$. 尤其是 $p(v) = E[a'uv + b'v1_{\mathcal{M}}]$, $p(u) = E[a'u^2 + b'u1_{\mathcal{M}}]$. 因此，$a', b'$ 可从下列方程解得 (注意：$E[u] = E[v] = 1$, $E[uv] = E[u]E[v] = 1$)：

$$p(v) = a' + b', \qquad p(u) = a'E[u^2] + b'.$$

这个方程组的行列式为 $1 - E[u^2] = E^2[u] - E[u^2] = \mathrm{Var}[u] \neq 0$. 因此，方程组有解。同时，因为 $E[r_u] \neq E[r_v]$ 和 $E[u] = E[v] = 1$ 可导得 $p(u) \neq p(v)$, 所以这样解得的 $a' \neq 0$. 而把这两个等式代入上式，我们立即得到

$$a' + b' - p(x) = \frac{E[ux]-1}{E[u^2]-1}(a' + b' - a'E[u^2] - b').$$

因此，(注意：$E[x] = 1$)

$$p(x) = a'E[ux] + b' = E[(a'u + b')x] = E[mx].$$

注意到 p 是齐次函数，这个等式对所有满足 $p(x) \neq 0$ 的 $x \in \mathcal{M}$ 成立。再由函数的连续性，它也对所有 $x \in \mathcal{M}$ 成立。再加上，$p(u) = E[mu] \neq$

$p(v) = E[mv]$ 和 $E[u] = E[v] = 1$ 可导出 m 与 $1_{\mathcal{M}}$ 一定不共线，而导出 $E[r_m] \neq E[1_{\mathcal{M}}]$, 因此，条件 1) 成立。

此外，不难验证，如果 $m \in \mathcal{M}$ 满足条件 1), 那么对于任何常数 $a \neq 0$ 和 b, $u = am + b$ 都有条件 2) 成立 (习题)。反之，如果 u 满足条件 2)，那么又必然有满足条件 1) 的唯一的 $m = a'u + b'$, $a' \neq 0$. 这就证明了定理中最后的论断。 □

这样，我们就导出了线性定价法则的等价形式：资本资产定价模型 (CAPM)。如所周知，资本资产定价模型是 Sharpe ([9]) 等人在 1960 年代提出的。这一模型告诉人们，每一种金融资产 (未定权益) 的期望超额收益 $E[r] - r_f$ 都与某种"市场组合"的期望超额收益 $E[r_u] - r_f$ 之间有线性关系。这就使人们对金融市场有一个整体的观念。因此，当它问世时，在金融学术界和业界都引起了很大的轰动，并且被称为"β 革命" (因为人们经常用 β 来表示这一线性关系中的系数)。后来，Black [1] 还针对没有无风险证券的市场改进了资本资产定价模型，而提出更一般的"零-β 资本资产定价模型"。

严格地说，尤其从经济意义上来说，这一 CAPM 并非是 Sharpe 等提出的 CAPM，因为这里并没有说出"市场组合" u 是什么。以后我们将从其他角度来讨论 CAPM，并指出"市场组合" u 的进一步含义。在以后的进一步的框架下的讨论中，我们可以指出，u 是市场中所有风险证券所形成的组合。不过在我们目前的讨论中，除了必要的数学框架和线性定价法则以外，没有任何其他假设。尤其是没有关于市场上的证券量的任何假设。因此，在目前的模型中不可能有"市场组合"的概念。至于定理 2.3 的结果则起源于 Ross [8] 和 Dybvig and Ingersoll [4]。

在结束本章以前，我们还可注意到下列数学命题：

命题 设 x 和 y 是两个随机变量。$\beta = \mathrm{Cov}[x, y]/\mathrm{Var}[x]$. 那么

$$\mathrm{Var}[y - \beta x] = \min_{\gamma \in \mathbb{R}} \mathrm{Var}[y - \gamma x],$$

即 βx 使所有 x 的倍数 γx 中与 y 之差的方差最小。

证明 事实上，

$$\mathrm{Var}[y - \gamma x] = \mathrm{Var}[y] - 2\gamma \mathrm{Cov}[x, y] + \gamma^2 \mathrm{Var}[x].$$

右端对 γ 求导，由极值的必要条件 (或利用二次函数的性质)，立即可得当 $\gamma = \beta = \mathrm{Cov}[x, y]/\mathrm{Var}[x]$ 时，$\mathrm{Var}[y - \gamma x]$ 达到极小。 □

由这条命题可知，上述的资本资产定价模型还有这样的统计意义：$\beta_{r_y r_u} r_u = \mathrm{Cov}[r_y, r_u] r_u / \mathrm{Var}[r_u]$ 使所有 γr_u 中与 r_y 之差的方差达到最小。

这相当于说，如果我们对证券 (或更一般的未定权益) y 的收益率 r_y 关于 r_u 作回归分析，即用最小二乘法，求 γ 和 α, 使 $\mathrm{Var}(r_y-\gamma r_u-\alpha)$ 的方差最小，那么 γ 的最优估计就是 $\beta_{r_y r_m}$, 而 α 的最优估计为 $r_f(1-\beta_{r_y r_m})$. 这说明，$\beta_{r_y r_m}$ 的表达式有其特殊含义，它意味着对证券收益率和“市场收益率”的实际数据应用 CAPM，与通常的回归分析在运算上是一样的。

说明 正如我们在“代引言”中所述，无套利假设首先是在 Modigliani–Miller 1958 年的经典论文中明确提出的。在目前流行的文献中，无套利假设通常采用 Duffie [3], Ingersoll [6] 的表达形式，其中把套利机会区分为第一种套利机会 (当前价值非正，未来价值为正) 与第二种套利机会 (当前价值为负，未来价值非负)。这样一来，数学上会增加很多不必要的麻烦。我们采用类似 Jarrow [7] 中对无套利假设的表达形式，其中明确提出“定价函数”的概念，并且通过它和假设 $E[m]\neq 0$ 已经把部分第一种套利机会排除掉，避免了一些烦琐讨论。同时，我们把无套利假设分成“五个层次”，这使得不同层次的无套利假设的含义更加明确。

明确以二期模型的形式的来讨论定价问题出现在 Ross 1978 年发表的 [8] 中。在那里， Ross 指出完整的无套利假设等价于定价函数是正线性函数。这一定理后来被称为资产定价基本定理。我们将在后面来讨论这一在数学上也很深刻的定理。但是对于金融经济学的大部分讨论，我们实际上只需要较简单的线性定价法则。这一思想也来自 Ross [8]，它与资本资产定价模型的联系则是 Dybvig–Ingsoll [4] 提出的。明确提出它们之间的等价性则出于 Cochrane [2]。不过我们的讨论有其自有的特色。Cochrane 的框架中，线性定价法则被假定为始终成立，而我们则从“正齐次定价法则”出发来讨论，证明是独特的，并且避免了一些不必要的繁复计算。

早期的金融经济学教科书都没有注意到线性定价法则与 CAPM 的等价性。但是 Duffie [3] 则把资产定价基本定理放在首要地位，并指出由它能导出 CAPM。请参看它的第 1 章。Ingersoll [6] 的第 2 章也是值得一读的材料。Huang–Lizenberger [5] 的有关材料在第 8 章，这是因为该书强调的是传统的“均衡定价”。令人奇怪的是 Jarrow [7] 中居然没有类似讨论。

思考题和习题

1. 试举出一些金融资产定价问题的实际例子。它们能否用我们的基本模型来估值？
2. 如果在现实中线性定价法则成立，那么所有银行的利率都应该是一样的。但是实际情况并非如此。怎样解释这种现象？
3. 怎样利用向量空间工具，来表达、理解无套利假设的各种层次，并解释其经济意义。
4. 什么是完全市场？它反映了一种怎样的现实金融市场的近似？它的基本性质是什么？

5. 同样是“不确定环境下的商品定价”，一种赌博游戏 (例如，“吃角子机”) 的“定价”与金融资产定价有何根本不同？

6. 什么是未定权益的 Hilbert 空间和线性定价法则的随机折现因子？

7. 在上一讲中讨论的简单的期权定价模型中，什么是该模型中的未定权益 Hilbert 空间和随机折现因子？如果未来有 S 种状态，那么未定权益 Hilbert 空间和随机折现因子又是什么？

8. 什么是未定权益的时间价值和风险价值？它们是相对什么而言的？

9. 怎样定义无风险证券和无风险利率？怎样理解无风险证券的模仿组合？

10. 设 $\mathcal{M}$ 是包含无风险证券的未定权益空间，即 $1 \in \mathcal{M}$. $\mathcal{M}_1 \subset \mathcal{M}$ 为 $\mathcal{M}$ 的子空间，但 $1 \notin \mathcal{M}_1$. 这时，从 $\mathcal{M}$ 出发来看问题，1 与 $1_{\mathcal{M}_1}$ 之间有什么关系？这种关系有什么经济意义？

11. 把上题中的无风险证券 1 换为随机折现因子 m, 讨论同样的问题。

12. 怎样来理解线性定价法则和资本资产定价模型的等价性？

13. 把资本资产定价模型中的 $r \in R_1$ 换为 $r \in R_0$, 即把模型中的“收益”换为“超额收益”，那么模型该怎么来叙述？它与线性定价法则又是什么关系？

14. 证明，如果 $m \in \mathcal{M}$ 满足 CAPM (定理 2.3 的条件 2)，那么对于任何常数 $a \neq 0$ 和 b, $u = am + b1_{\mathcal{M}}$ 也满足 CAPM。

15. 证明，对于两个方差有限数学期望不为零的随机变量 x 和 y, 如果

$$E[x] = \frac{\mathrm{Cov}[x,y]}{\mathrm{Var}[y]}E[y], \qquad E[y] = \frac{\mathrm{Cov}[x,y]}{\mathrm{Var}[x]}E[x],$$

那么存在常数 $a \neq 0$ 和 b, 使得 $y = ax + b$. 把这个结果用到 CAPM 的讨论上，你能得到什么结果？

参考文献

[1] Black, F., 1972, Capital market equilibrium with restricted borrowing, *Journal of Business,* **45:** 444–454.

[**2**] Cochrane, J. H., 2001, Asset Pricing, Princeton University Press, Princeton, N. J..

[**3**] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N.J..

[**4**] Dybvig, P. H. & J. E. Jr. Ingersoll, 1982, Mean-variance theory in complete markets, *Journal of Business*, **55:** 233–251.

[**5**] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[**6**] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[**7**] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[**8**] Ross, S. A., 1978, A simple approach to the valuation of risky streams, *Journal of Business*, 51: 453–475.

[**9**] Sharpe, W., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, **19:** 425–442.

附录：数学预备知识 1

1. 实数域 $\mathbb{R}$**：** $\mathbb{R}$ 用来表示实数全体。实数是通过： [自然数 $\Rightarrow$ 整数 $\Rightarrow$ 有理数 $\Rightarrow$ 实数] 这样的过程逐步扩充而成的。自然数全体似乎是人类“先验”就有的。但在理论上，它同样需要用“公理”来定义。有关的公理称为“Peano 公理”。这里我们没有必要去细究。以后的扩充首先是为使各种代数运算能进行而形成的。从自然数到整数是因为要使减法总有可能进行；从整数到有理数是因为要使除法总有可能进行。有理数全体常称为**有理数域**。“域”就是指其中能进行加减乘除四则运算。从有理数扩充到实数比较复杂。中学数学中称实数为有理数与无理数全体，而无理数则定义为“无限不循环小数”。其实无理数是通过有理数的极限来定义的。为使定义极限有可能，这里也需要一条公理，它称为“连续性公理”。在实数范围内也同样能进行四则运算，因此，常常也称它为“实数域”。而实数域比有理数域更有用之处在于其中能比较方便地运用极限概念。

经济学中为什么需要用实数域来描述？细究起来好象也没有多大道理。实际经济生活中用有理数已足以刻划。但是为了研究与极限概

念有关的趋向问题，理论上必须采用实数域以及由此产生的微分学等等才能说得清楚。对此，似乎也只能采取 Friedman 工具主义的态度。

2. 函数与映射：一个实变量实值函数表示为 $f{:}\mathbb{R}\to\mathbb{R}$, 它表示一个自变量为实数、因变量也是实数的函数，即它对每一个实数 (自变量) $x\in\mathbb{R}$ 指定了另一个实数 (因变量) $y=f(x)\in\mathbb{R}$ 与它相对应。比函数更一般的概念是映射。它可以对任意两个集合 A 和 B 来定义。映射 $g{:}A\to B$ 是指对每一个元素 $x\in A$ 定义了另一个元素 $y=g(x)\in B$. 例如，这里 A 和 B 都是实数域的子集，那么 g 就是更一般的实变量实值的函数。在我们的讨论中，我们经常需要 A 或 B 是向量空间 $\mathbb{R}^n$ 或者其子集的情形。

3. 向量空间和矩阵：把若干个实数按一定方式排列起来，它们的全体就形成一个向量空间。通常一个 n 维向量空间是指 n 个按次序排列的实数 y_i $(i=1,2,\ldots,n)$ 所组成的数组 $y=(y_1,y_2,\ldots,y_n)$ 全体。它常常记为 $\mathbb{R}^n$. 而 $\mathbb{R}$ 中的元素则称为 n 维向量。在一个向量空间中可以进行加减运算，但不能进行乘除运算。然而，一个 n 维向量与一个实数是可以相乘的。向量空间的严格的公理化就是用能进行这两种运算来定义的。由一个 1 和 $n-1$ 个 0 所构成的 n 维向量称为 $\mathbb{R}^n$ 的**单位向量**。这样的向量一共有 n 个。每个 n 维向量都可表示为它们的**线性组合**，即它们乘上实数后的和。如果有一组向量也有这样的性质，那么这一向量组就称为 $\mathbb{R}^n$ 的**基**。一组向量基中的向量个数一定不少于 n 个。否则一定有某个向量不能表示为这组向量的线性组合，从而这组向量不能形成基。如果我们讨论的空间是证券组合的未来 S 个状态下的价值所形成的 S 维空间，那么 S 个单位向量就是所谓 Arrow–Debreu 证券。如果基本证券组形成这个 S 维空间的基，那么所对应的证券市场就是完全的；否则就是不完全的。在完全市场中，每一未定权益 (即 S 维向量) 都可用基本证券的线性组合来表示，从而未定权益就可用基本证券组来定价。而在不完全市场中，就存在某些未定权益不能表示为基本证券的线性组合，从而用基本证券来定价就有麻烦。但是通过完整的无套利假设，可用基本证券组来为它定出一个价格区间。

一般情况下，我们不一定把 n 个实数这样横排起来。为了某种需要，我们也可以把它竖排起来。在后一种情形下，我们记 $x=(x_1,x_2,\ldots,x_n)^T$, 这里 T 就表示把这 n 个实数竖排起来，即所谓“转置”。我们称这种形式的向量为**列向量**, 而前面形式的向量称为**行向量**。n 维列向量全体与 n 维行向量全体都形成 n 维向量空间，但是有时有必要把它们区分为两个不同的向量空间。对一个行向量 $\theta=(\theta_1,\theta_2,\ldots,\theta_n)$ 与一个列向量 $x=(x_1,x_2,\ldots,x_n)^T$ 可定义它们的 (不可交换的) 乘积，即

$$\theta\cdot x=\theta_1x_1+\theta_12x_2+\cdots+\theta_nx_n\in\mathbb{R}.$$

在我们前面的讨论中，行向量 θ 的意义是证券组合，列向量 x 的意义

是证券价格，它们的乘积的意义就是证券组合的总价值。这时，把行向量与列向量区别开来，在经济含义上就很有好处。不过在数学上也经常不加区分。这时就没有行向量与列向量的区别。而相应的乘积则称为两个向量的**内积**。

对于一个 n 维向量空间 $\mathbb{R}^n$ 来说，内积是一个非常重要的概念。它是反映两个向量之间的夹角的量度。定义了内积的 n 维向量空间称为 n 维**欧几里德空间**或简称**欧氏空间**。对于 $\mathbb{R}^n$ 中的任何两个向量 $x=(x_1,\ldots,x_n)$ 和 $y=(y_1,y_2,\ldots,y_n)$, 其内积常表示为

$$(x,y)=x_1y_1+x_2y_2+\cdots+x_ny_n.$$

$$\|x\|=\sqrt{(x,x)}=(x_1^2+x_2^2+\cdots+x_n^2)^{1/2}$$

称为向量 x 的**长度**。$\|x-y\|$ 称为向量 x 与 y 之间的**距离**。而它们之间的夹角 α 则由下式定义：

$$\cos\alpha=\frac{(x,y)}{\|x\|\|y\|}.$$

尤其是 $(x,y)=0$ 称为 x 与 y **垂直**或**正交**。在 $\mathbb{R}$ 中有了距离以后，我们就可在其中定义**极限**概念。也就是说，$\mathbb{R}^n$ 中的一个点 (向量) 列 $\{x(k)\}$ 趋向于一个固定点 (向量) a, 是指 $\|x(k)-a\|\to 0$. 由此还可定义闭集与开集的概念。$\mathbb{R}^n$ 中的一个集合 A 称为**闭集**是指其中每一个有极限的点列的极限也在其中。**开集**则可定义为一个闭集的余集。

4. 欧氏空间和 Hilbert 空间：我们在前面对 n 维向量空间已经定义过内积的概念。定义了内积的 n 维向量空间称为欧氏空间。假设这个欧氏空间为 X, (x,y) 表示 $x,y\in X$ 之间的内积。那么不难验证它有以下几条性质：对于任何 $x,y,z\in X$,

1. $(x,y)=(y,x)$.
2. 对于任何实数 $\lambda,\mu\in\mathbb{R}$, $(\lambda x+\mu y,z)=\lambda(x,z)+\mu(y,z)$.
3. $(x,x)\geqslant 0$, 并且 $(x,x)=0$ 当且仅当 $x=0$.

这三条性质与 X 的维数没有关系。因此，它可以推广到一般的情形。事实上，设 X 是一般的向量空间，即其中的元素可相加以及与任意实数相乘。如果在 X 上定义一个用 (x,y) 来表示的二变量函数，满足上述三条性质，那么 X 就称为**内积空间**。与欧氏空间一样，我们同样定义 x 的长度 $\|x\|=\sqrt{(x,x)}$. 不过经常把 $\|x\|$ 称为 x 的**范数**。这样，如果有一个集合 (例如，$\mathcal{M}$ 为一些方差有限的随机变量所构成的向量空间)，其

上可定义一种内积 (例如，对 $\mathcal{M}$ 中的随机变量 x, y 定义 $(x, y) = E[xy]$)，那么欧氏空间的许多结论都可对这样的内积空间成立 (对于 $\mathcal{M}$ 来说，$E[x^2] = 0$ 导出 $x = 0$ 是意味着 x 作为随机变量，以概率 1 等于 0. 同时，$\mathcal{M}$ 中任何两个随机变量相等也都是意味着它们相等的概率等于 1, 不相等的概率等于 0)。不过欧氏空间有一条称为**完备性**的性质还没有包含在内：*如果序列* $\{x_n\} \subset X$ *满足* $\lim_{n,m\to\infty} \|x_n - x_m\| = 0$, *那么存在* $\bar{x} \in X$, *使得* $\lim_{n\to\infty} \|x_n - \bar{x}\| = 0$. 由于我们现在考虑的向量空间可能是无限维的，它不一定有这条很有用的性质。把这条性质加到一个内积空间中去，使它更接近欧氏空间，这就定义了所谓 **Hilbert 空间**，即它是完备的内积空间。

所有方差有限的随机变量所形成的集合，如果用 $(x, y) = E[xy]$ 来定义其上的内积，那么它就是一个 Hilbert 空间。通常记作 $L^2(P)$, 其中 P 表示概率测度。有限个方差有限的随机变量张成的向量空间，通过同样的办法来定义内积，它也是一个 Hilbert 空间。但它在本质上与欧氏空间没有多大两样。无限多个方差有限的随机变量张成的内积空间不一定是完备的。如果它是完备的，那么它是 Hilbert 空间，并且就是 $L^2(P)$ 中的一个闭子空间；反之，$L^2(P)$ 的任何一个闭子空间也可看作若干方差有限的随机变量所张成的 Hilbert 空间。在我们的讨论中，一般总假定未定权益空间 $\mathcal{M}$ 是 $L^2(P)$ 的闭子空间，即它也是 Hilbert 空间。

5. Hilbert 空间中的正交性：对一个 Hilbert 空间中的元素来说，如果 $(x, y) = 0$, 那么与以前一样，我们称 x 与 y 正交。两个集合互相正交则是指一个集合中的任何一个向量与另一个集合中的任何一个向量都正交。对于正交性，我们有下列正交分解定理成立：

正交分解定理　*设 X 为一个 Hilbert 空间。$Y \subset X$ 为它的一个闭子空间，并且 $Y \neq X$. 那么 X 一定可以分解为 Y 与另一个与 Y 正交的线性子空间 $Y^\perp$ 的直和:*

$$X = Y \oplus Y^\perp,$$

即对于任何 $x \in X$, 存在唯一的 $y_x \in Y$ 和 $y_x^\perp \in Y^\perp$, 使得 $x = y_x + y_x^\perp$.

证明概要　在 Y 中求一个与 x 的距离最近的向量 y_x, 即 $y_x \in Y$ 满足 $\|x - y_x\| = \min_{y\in Y} \|x - y\| = d$. 事实上，我们可以找到一个点列 $\{y_x^n\} \subset Y$, 使得 $\|x - y_x^n\| \to d$. 我们指出 $\lim_{n,m\to\infty} \|y_x^n - y_x^m\| = 0$. 首先，我们可直接验证下列“平行四边形法则”成立：

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

从而当 $n, m \to \infty$ 时，

$$\|y_x^n - y_x^m\|^2 \;=\; \|(x - y_x^n) - (x - y_x^m)\|^2$$

$$
\begin{aligned}
&= 2(\|x-y_x^n\|^2+\|x-y_x^m\|^2)-\left\|2\left(x-\frac{y_x^n+y_x^m}{2}\right)\right\|^2 \\
&\to 2(d^2+d^2)-4d^2=0
\end{aligned}
$$

因此， y_x^n 有极限 y_x. 由于 Y 是 X 的闭子空间，故 $y_x \in Y$.

另一方面，令 $y_x^\perp = x - y_x$, 我们指出，$(y_x, y_x^\perp) = 0$. 事实上，对于任何 $y \in Y$ 有 $\|x-y_x\| \leqslant \|x-y\|$, 从而对于任何 $y \in Y$ 和任何 $t > 0$, 有

$$
\begin{aligned}
\|x-y_x\|^2 &= (x-y_x, x-y_x) \\
&\leqslant \|x-y_x-t(y-y_x)\|^2 \\
&= (x-y_x-t(y-y_x), x-y_x-t(y-y_x)) \\
&= (x-y_x, x-y_x)-2t(x-y_x, y-y_x)+t^2(y-y_x, y-y_x)
\end{aligned}
$$

因此，对于任何 $y \in Y$ 和 $t > 0$ 有

$$
(x-y_x, y-y_x) \leqslant \frac{t}{2}\|y-y_x\|^2.
$$

因为 t 可任意接近零，$y - y_x$ 又可取为 Y 中的任何元素，所以 $x - y_x$ 必定与 Y 中的任何元素正交，即 $y_x^\perp = x - y_x \in Y^\perp$. □

这一正交分解非常有用。以后，我们将用它来讨论线性方程组。在本讲中，我们将拿它来证明 Riesz 表示定理。

6. Hilbert 空间中的 Riesz 表示定理： Riesz 表示定理指出，Hilbert 空间中的连续线性函数一定可通过某个元素对其他元素的内积来表示。它对金融经济学的意义在于：如果“市场”(由方差有限的某些随机变量 (证券的未来价值) 所张成的 Hilbert 空间) 有连续的线性定价函数，那么它一定可通过某个“定价证券” (即“随机折现因子”) 来表示。

Riesz 表示定理 设 X 是一个 Hilbert 空间。如果 $f\colon X \to \mathbb{R}$ 是一个连续线性函数，那么存在唯一的 $x_f \in X$, 使得对于任何 $x \in X$, 有 $f(x) = (x_f, x)$.

证明概要 设 $Y = \{x \in X \mid f(x) = 0\}$. 那么不难验证，$Y$ 是 X 的闭子空间。任取非零的 $z \in Y^\perp$, 定义 $x_f = zf(z)/\|z\|^2$, 不难验证，x_f 就是唯一的满足定理要求的元素。 □

三、公司财务的 Modigliani–Miller 定理

本讲要求：利用上一讲中提出的无套利假设线性定价法则，讨论公司财务的 Modigliani–Miller 定理。理解所有有关 MM 定理的讨论无非是线性定价法则的应用。

数学预备知识：随机变量的概念。初等代数运算。

内容：

- 3.1 Modigliani–Miller 定理与线性定价法则。
- 3.2 关于分红政策的 Modigliani–Miller 定理。
- 3.3 关于资本结构的 Modigliani–Miller 定理。

3.1 1958 年起，Modigliani 和 Miller 发表了一系列论文，探讨“公司的财务政策是否会影响公司的价值”这一主题。这里的财务政策是指分红政策、资本结构等。他们的结论是：在理想的市场条件下，公司的价值与这些政策无关。后来他们的这些结论就被称为 Modigliani–Miller 定理 (MMT)，并且为公司财务这门学科奠定了基础。

上一讲中讨论的无套利假设最早就出现在这些经典论文中。当时的无套利假设在数学上还不能表达得很清楚，因而他们的推理也显得比较累赘。今天我们已经可以把无套利假设用一个定价函数来表达，这就可以使推导大为简化。实际上，Modigliani–Miller 定理要求的无套利假设仅仅是线性定价法则。而这一法则其实就是 Modigliani 和 Miller 的讨论的本质。换句话说，他们企图得到的结果，就是：“未来”价值一样的“未定权益”(以及所有与它们有同样外延的对象，尤其是包括他们所需要的“公司价值”)，其“当前”价值也一样，而与它们的“当前”组成成分 (资本结构等等) 无关。对于这一定理，有一种很形象的说法为：一个蛋糕的大小与切蛋糕的方式是无关的。

在这一讲中，我们就把 Modigliani–Miller 定理作为线性定价法则的最简单的应用来进行讨论。尽管在历史上，Modigliani–Miller 定理其实比资产定价基本定理和明确的线性定价法则早出现 20 年。

现在我们来进入数学形式讨论。我们的模型仍然与以前一样是一个二期模型，即有当前和未来两个时刻。市场中涉及的金融资产或未定权益的未来价值是随机变量，而它们涉及的当前价值则是确定的，并且未来价值与当前价值之间的关系由一线性定价函数来决定。由于在这一讲中，我们考虑的对象不一定是证券或证券组合的价值，而更多的是公司的价值，因此在这里需要用一个更好的专门名词来称呼它们。这一名词就是所谓“现金流 (cash flow)”。现金流原来的含义是指

某一对象随时间变化的现金价值。现在在这里则是指对象的未来现金价值。它相当于我们前面所说的未定权益的未来价值，是一个随机变量。也就是说，随机变量现在在这里又多了一个外延上相同的名词。与以前一样，读者不要从“现金流”的内涵上去理解它在这里的含义。

另一方面，本讲的模型不涉及随机折现因子，因此，我们的数学框架可仅仅假设为“现金流随机变量全体构成的一个线性空间 $\mathcal{C}$ 以及其上定义的线性定价函数 $p{:}\,\mathcal{C}\to\mathbb{R}$。”同时，我们也不一定假设 $\mathcal{C}$ 中有“无风险现金流”，即常数 $1\in\mathcal{C}$. 当需要假设时，我们将专门作说明。这时，我们通常假定 $p(1)\geqslant 1$. 正如我们在上一讲中所指出的，$r_f=1/p(1)$ 的经济含义就是无风险利率。但在本讲中我们将只使用记号 $p(1)$, 而不用 r_f. 至于 Hilbert 空间之类的要求在本讲中则不需要。

3.2　我们先讨论分红政策。考虑两个在当前有同样的股份数 n、而未来有同样的债务、投资以及同样市值的公司 A 和 B. 我们的问题是：如果这两个公司采取不同的分红政策，那么它们的当前价值是否会不同？为此，设 $d_A,d_B\in\mathcal{C}$ 分别为它们在未来的每股的分红值。它们都是“现金流随机变量”。 $s_A,s_B\in\mathcal{C}$ 分别为它们在未来的股票价格。它们也是“现金流随机变量”。于是如果有人在未来持有 A 或 B 的一份股票，那么该股票在未来的价值应该是 d_A+s_A 或 d_B+s_B. 而 A 或 B 的股票在当前的价格则应是 $p(d_A+s_A)$ 和 $p(d_B+s_B)$. 由于分红政策的不同，在未来的股份数会有所不同。我们设 Δn_A 和 Δn_B 分别为它们在未来的股份变动数。这里我们可假设它们也是随机变量。但它们并非是“现金流随机变量”，而只是与“现金流随机变量”处于同一概率空间的随机变量。我们的结论如下：

定理 3.1 **(Modigliani–Miller 定理之一)**　*假设在现金流随机变量线性空间 $\mathcal{C}$ 中存在线性定价函数 p. 如果公司 A 和 B 在当前有相同的确定股份数 n, 在未来有相同的* (随机) *市值，即*

$$(n+\Delta n_A)s_A=(n+\Delta n_B)s_B, \tag{1}$$

以及有与分红政策无关的相同的债务－投资支出，即

$$nd_A-\Delta n_A s_A=nd_B-\Delta n_B s_B, \tag{2}$$

那么 A 和 B 在当前有相同的股票价格，即

$$p(d_A+s_A)=p(d_B+s_B). \tag{3}$$

证明　由 (1) + (2) 即得

$$n(d_A+s_A)=n(d_B+s_B).$$

因此，$d_A + s_A = d_B + s_B$. 从而 (3) 成立。 □

这条定理在数学上只需用一些简单的代数运算，即使把未来的不确定性考虑在内，也不牵涉到复杂的推导。问题在于其中的经济内容如何来理解。(1) 式的含义很清楚，其两端就是公司股份市值的定义。费解的是 (2) 式，为什么它意味着两个公司有相同的债务－投资支出。事实上，等式的两端可看作公司支付了它们的除分红以外的债务和所有的投资支出以后的余额现金流；这些余额将用来支付分红，对于两个公司来说，即分别是 nd_A 和 nd_B，以及支付回购股票，即分别是 $-\Delta n_A s_A$ 和 $-\Delta n_B s_B$ (严格来说，这都是指 Δn_A 和 Δn_B 为负数的情形；如果它们是正数，相应的两项应该是增股得到的直接融资的相反值)。因此，(2) 式的经济含义应该是两个公司有相同的债务－投资支出。这样，定理 3.1 说明的是：两个公司除了未来 (不确定的) 分红政策和 (不确定的) 股价以外，其他方面全部相同。结论是两个公司当前的股价是相等的。

这就是著名的 Modigliani–Miller 定理的一个结论。不过它出现在 Modigliani–Miller 1961 年的论文 [10] 中。我们应该注意到这里对定理的成立有许多条件。如果这些条件与现实情况出入不大，那么我们就应该在现实世界中观察到各公司的分红政策是相当随意的。然而，在现实世界中并未观察到这样的现象。事实上，各个公司的分红政策是相当稳定的。如果不发生意外，一般不会改变。因此，这一定是定理 3.1 中有的假设与实际不符。后来有许多研究就致力于究竟是什么原因导致这一结果不成立。这里可找的原因也就是 (1) 与 (2). 当然，还有就是线性定价法则，但这一法则至少作为理论的简化应该成立。下面我们分别来讨论 (1) 和 (2) 不成立的原因。

(1) 实际上意味着分红政策不影响公司的市值，尤其是不影响股价。而这是与事实不符的。这是因为在股市上，信息并不完全透明。于是当分红提高时，人们会以为这是利好消息，而促使人们预期股价上升，以至预期公司的市值上升。其最终结果显然是预期分红好的公司当前的股价要高。换句话说，(1) 的成立需要信息完全透明，即不存在所谓“信息不对称”的现象。在金融经济学的经典讨论中，信息的作用很少进入理论框架。而这里不涉及信息的作用是得不到进一步的结果的。因此，要改进 (1) 就必须在进一步的理论框架中进行。这超出了我们的讲义的范围。

(2) 不成立的原因之一是其中没有考虑个人所得税问题。如果我们考虑分红所得的税率与资本增益所得的税率有所不同，那么结论也就会有变化。这是在我们的框架中可以讨论的问题。

现在假定所有投资者面临同样的税率，分红所得税率为 τ_d, 资本增

益所得税率为 τ_c, 且它们满足

$$0 \leqslant \tau_c < \tau_d < 1. \tag{4}$$

在这种情况下，对于公司 A 的一份股票来说，它在未来的税前现金流随机变量为 $d_A + s_A$; 而其税后现金流随机变量 y_A 将满足方程

$$y_A = d_A(1-\tau_d) + [s_A - p(y_A)](1-\tau_c) + p(y_A), \tag{5}$$

其中 $p(y_A)$ 是 A 公司的当前股价 (注意我们前面提到的关于无风险现金流与常数不加区别的说明，这里 $p(y_A)$ 其实是一项无风险权益)。对于 B 也有类似的方程。于是我们又有下列定理：

定理 3.2　在定理 3.1 的假设下，如果分红所得税率为 τ_d, 资本增益所得税率为 τ_c, 且它们满足 (4), $d_A > d_B$ (它意味着 $d_A \leqslant D_B$ 的概率为零，但 $d_A > d_B$ 的概率大于零)，并且无风险现金流 $1 \in \mathcal{C}$, 其当前价值 $p(1) \leqslant 1$. 还假定 p 是正线性函数，即当 $d_A > d_B$ 时, $p(d_A) > p(d_B)$. 那么 A 和 B 在当前的股票价格 $p(y_A)$ 和 $p(y_B)$ 满足

$$p(y_A) < p(y_B). \tag{6}$$

即未来分红较高的公司当前的股价较低。

证明　事实上，由 (5) 可得

$$\begin{aligned} p(y_A) &= p(d_A(1-\tau_d) + [s_A - p(y_A)](1-\tau_c) + p(y_A)) \\ &= p[(1-\tau_c)(d_A+s_A) - (\tau_d-\tau_c)d_A] + \tau_c p(y_A)p(1), \end{aligned}$$

其中用到 $p(p(y_A)) = p(y_A)p(1)$. 由此可解得

$$p(y_A) = \frac{p[(1-\tau_c)(d_A+s_A) - (\tau_d-\tau_c)d_A]}{1-\tau_c p(1)}.$$

同理，

$$p(y_B) = \frac{p[(1-\tau_c)(d_B+s_B) - (\tau_d-\tau_c)d_B]}{1-\tau_c p(1)}.$$

由此可得，

$$p(y_B) - p(y_A) = \frac{(\tau_d-\tau_c)p(d_A-d_B)}{1-\tau_c p(1)} > 0.$$

□

这一定理说明，为在所得税上占点便宜和抬高当前的股价，在 $p(1) \leqslant 1$ 的假设下 (这一假设显然也起着关键作用)，分红应该越少越好，最好

是不分红。这点首先与前面说的“预期分红高的股价高”的现象不符。当然，这是从股价的不同侧面来考虑问题的缘故，“预期分红高的股价高”往往是由股市心理引起的，它似乎应该属于行为金融学研究范畴。不属于这里的研究框架。但是即使不考虑这一因素，仍然可以发现，这同样与实际不符。有人把它解释为，归根到底，所有公司面临同样的税率这一假定与实际不符。Miller 和 Scholes 在文献 [8] 中对分红与税收之间的关系进行了专门的研究。他们的结论是公司会采取种种措施使得分红不纳所得税，以至实际上 $\tau_d \approx \tau_c$. 这样又将导至与实际不符的股价与分红政策无关的结论。这说明模型仍然有不完善之处。可能最根本的还是两个各方面完全一样的公司是不存在的。总之，这方面的研究停留在这一层次上是远远不够的。

3.3　另一条 Modigliani–Miller 定理是说公司的价值与公司的资本结构无关。一般情况下，提到 Modigliani–Miller 定理或命题，都是指这一条定理。这里的资本结构是指公司的债权与股权之间的比例。在他们的研究以前，曾经有一些经济学家认为，应该存在一个最佳的债权/股权比，使得公司的价值最高。而 Modigliani 和 Miller 则通过无套利假设导出的线性定价法则证明了公司的价值是与此无关的。对此，我们仍可通过前面的模型来讨论。

假定有公司 A 和 B, 公司 A 既发行股票，又发行债券；而公司 B 则只发行股票，不发行债券。这两个公司在未来的价值 $x_A, x_B \in \mathcal{C}$ 是一样的，即 $x_A = x_B$. 这里要注意，这个等式意味着两个随机变量相等。但是

$$x_A = n_A x_A^e + m_A x_A^b, \qquad x_B = n_B x_B^e, \tag{7}$$

其中 $x_A^e \in \mathcal{C}$ 是 A 公司未来的股价，它是现金流随机变量；n_A 是 A 公司的确定股份数；$x_A^b \in \mathcal{C}$ 是 A 公司未来的债券价。由于公司有可能破产，债券价格有时也可能是随机变量，因而它也是 $\mathcal{C}$ 的元素；m_A 是公司 A 发行的确定债券数；$x_B^e \in \mathcal{C}$ 和 n_B 是对于公司 B 的未来的 (随机) 股价和 (确定) 股份数。

定理 3.3 **(Modigliani–Miller 定理 1)**　假设在现金流随机变量线性空间 $\mathcal{C}$ 中存在线性定价函数 p. 如果公司 A 和 B 的未来价值相等，即 $x_A = x_B$, 而它们的资本结构分别满足 (7), 那么它们的当前价值 V_A 和 V_B 也相等。

证明　事实上，由定价函数 p 的线性性，我们有

$$\begin{aligned} V_A &= n_A p(x_A^e) + m_A p(x_A^b) \\ &= p(n_A x_A^e + m_A x_A^b) \\ &= p(x_A) = p(x_B) = p(n_B x_B^e) = n_B p(x_B^e) = V_B. \end{aligned}$$

□

令 r 为债券的 (总) 收益率。在线性定价法则下，A 和 B 两个公司发行的债券的收益率 (假定两种债券都无风险) 应该是一样的。因此，$r = x_A^b/p(x_A^b) = x_B^b/p(x_B^b)$. 另一方面，由定理 3.3 可知，两个公司的总价值的 (总) 收益率 ρ 也是一样的，即 $\rho = x_A/p(x_A) = x_B/p(x_B)$. 这样，对于公司 A 的股票的 (总) 收益率 $i_A = x_A^e/p(x_A^e)$ 应该有下列推论成立：

推论 (**Modigliani–Miller 定理 2**) 在定理 3.3 的假设下，如果公司 A 和 B 的未来价值相等，即 $x_A = x_B$, 那么公司 A 的股票收益率 i_A 为其债权/股权比的线性函数，即

$$i_A = \rho + (\rho - r)\frac{D_A}{S_A}, \tag{8}$$

其中 $S_A = n_A p(x_A^e)$, $D_A = m_A p(x_A^b)$.

证明 事实上，

$$\begin{aligned} i_A &= \frac{n_A x_A^e}{n_A p(x_A^e)} = \frac{x_A - m_A x_A^b}{S_A} \\ &= \frac{\rho(D_A + S_A) - rD_A}{S_A}. \end{aligned}$$

因此，(8) 成立。 □

定理 3.3 及其推论 (它们正是最初的两条 Modigliani–Miller 定理) 同样在数学上是平凡的，并且我们从证明中可以看出，其中其实并不涉及股票、债券的具体性质。如果一个公司的价值可以表现为若干种证券的不同组合，那么定理的结论仍然成立。从这个角度来看，这条定理无非是说，两个证券组合在未来的值相同，那么它们在当前的值也相同。而这正是我们在本讲的一开始就说过的事，即它是线性定价法则的结果。不过当年 Modigliani 和 Miller 并未把他们的证明写得如此形式化，而是直接从 $V_A \neq V_B$ 来导出一个套利机会。其叙述自然要比上面麻烦。从这里也可以看出数学形式化的好处。

然而，如上叙述的 Modigliani–Miller 定理也与事实不符，即现实经济中的公司的债权/股权比有其一定规律。这就需要改进上述定理。改进从两方面来考虑。一是从公司所得税方面来考虑；一是从个人所得税方面来考虑。

设公司 A 发行的债券的当前价格为 $D = p(x_A^b)$, 债券规定的 (净) 利率为 i, 即如果公司不破产，它将对每份债券支付 $(1+i)D$. iD 将是每

份债券的利息。考虑公司面临的是类似美国所处的情况：债券利息支出可以抵税，即使公司面临破产也是如此。如果公司税率为 τ_c, 那么公司 A 的抵税额为 $i\tau_c m_A D$. 于是我们又有下列定理：

定理 3.4 在定理 3.3 的假设下，如果两公司税后的现金流分别为

$$x_A(1-\tau_c)+i\tau_c m_A D, \qquad x_B(1-\tau_c),$$

其中 i, τ_c, D 如上所述，并且无风险现金流 $1\in\mathcal{C}$, 其当前价值 $p(1)\leqslant 1$, 那么两公司的当前价值 V_A 和 V_B 满足下列等式:

$$V_A = V_B + i\tau_c m_A D p(1). \tag{9}$$

证明 事实上，

$$\begin{aligned} V_A &= p(x_A(1-\tau_c)+i\tau_c m_A D) \\ &= p(x_A)(1-\tau_c)+(i\tau_c m_A D)p(1) \\ &= p(x_B)(1-\tau_c)+(i\tau_c m_A D)p(1) \\ &= p(x_B(1-\tau_c))+(i\tau_c m_A D)p(1) \\ &= V_B+(i\tau_c m_A D)p(1). \end{aligned}$$

□

这条定理从债券利息可抵税的角度来考察两个未来价值一样的公司在当前的价值相差。结论竟是发行债券越多的公司的当前价值越大。这一结论似乎是说，在美国那样的税收政策下，直接融资不如间接融资。这同样是与事实不符的。问题在于借债太多，公司资不抵债时就要破产。而破产是要付出代价的。如果进一步考虑破产费用，那么改进后的定理就会变得比较合理。

为此，我们需要定义“破产费用”。这是一项未定权益，在不破产时它为零，而在破产时它大于零，并且是债券当前价格 D 的递增函数。具体地说，A 公司的破产费用是一项依赖于 A 的现金流的函数。它们之间的关系就像原生证券与衍生证券之间的关系。破产就是发生了资不抵债的状况。A 公司的“资”是它的税后现金流，如上所述，它等于 $x_A(1-\tau_c)+i\tau_c m_A D$; A 公司的“债”则是 $m_A D(1+i)$. 于是破产费用 c 应定义为

$$c=\begin{cases} 0, & \text{当 } x_A(1-\tau_c)+i\tau_c m_A D \geqslant m_A D(1+i), \\ c(D). & \text{当 } x_A(1-\tau_c)+i\tau_c m_A D < m_A D(1+i), \end{cases} \tag{10}$$

其中 $c(D)$ 表示它是债券当前价格 D 的递增函数。破产费用 $c \in \mathcal{C}$ 的当前价值自然就是 $p(c)$. 在这样的假定下，我们有下列定理：

定理 3.5　*在定理 3.3 的假设下，如果 A 公司的破产费用为如 (10) 所定义，那么公司 A 和 B 的当前价值 V_A 和 V_B 满足下列等式:*

$$V_A = V_B + i\tau_c m_A D p(1) - p(c), \tag{11}$$

$p(c)$ 为破产费用 c 的当前价值。

这一定理的推理除了 A 公司的价值中多了一项外，与上面几乎完全一样。其详细推导不妨从略。定理说明公司 A 必须在发行债券减税所得与破产费用之间作出权衡，而并非是债越多越好。这就与实际情况比较接近。

现在我们考虑从个人所得税方面来改进 Modigliani–Miller 定理 3.3. 假设两个公司的投资者都面临同样的个人所得税率 τ_b 和 τ_e, 其中 τ_b 是债权所得的所得税率，τ_e 是股权所得的所得税率；它们都是小于 1 的正数。同时假设公司 A 的债券是无风险的，即 A 满足下列不破产条件：

$$x_A(1-\tau_c) + i\tau_c m_A D \geqslant m_A D(1+i).$$

在这种情况下，公司 A 和公司 B 的税后现金流 y_A 和 y_B 应该分别满足下列方程：

$$y_A = (x_A - i m_A D)(1-\tau_c)(1-\tau_e) + i m_A D(1-\tau_b) + \tau_e p(y_A), \tag{12}$$

$$y_B = (x_B)(1-\tau_c)(1-\tau_e) + \tau_e p(y_B). \tag{13}$$

为指出这两个方程，对 A 公司分别考虑股权所有者全体和债权所有者全体的利益。对于股权所有者的税后现金流满足下列方程：

$$n_A Z_A^e = [(x_A - i m_A D)(1-\tau_c) - m_A D - n_A p(Z_A^e)](1-\tau_e) + n_A p(Z_A^e),$$

这里前一项为税后的资本增益收入，而后一项是股本值。债权所有者的税后现金流为

$$m_A Z_A^b = m_A D + i m_A D(1-\tau_b).$$

把两式相加，并注意到 $D = p(Z_A^b)$, $y_A = m_A Z_A^b + n_A Z_A^e$ 和 $p(y_A) = m_A p(Z_A^b) + n_A p(Z_A^e)$, 即得 (12). (13) 同理可得。有了这些讨论以后，我们就有下列定理：

定理 3.6　*在定理 3.3 的假设下，如果两个公司的投资者都面临同样的个人所得税率 τ_b 和 τ_e, 其中 τ_b 是债权所得的所得税率，τ_e 是股权*

所得的所得税率，那么公司 A 和 B 的当前价值 V_A 和 V_B 满足下列等式：

$$V_A = V_B + im_A D p(1)[(1-\tau_b) - (1-\tau_c)(1-\tau_e)]/(1-\tau_e p(1)). \tag{14}$$

证明 事实上，

$$\begin{aligned} V_A &= p(y_A) = p((x_A - im_A D)(1-\tau_c)(1-\tau_e) + im_A D(1-\tau_b) + \tau_e p(y_A)) \\ &= p(x_A)(1-\tau_c)(1-\tau_e) + im_A D[(1-\tau_b) - (1-\tau_c)(1-\tau_e)]p(1) \\ &\quad +\tau_e p(y_A)p(1). \end{aligned}$$

由此可解得

$$\begin{aligned} V_A &= p(y_A) \\ &= \{p(x_A)(1-\tau_c)(1-\tau_e) + im_A D[(1-\tau_b) - (1-\tau_c)(1-\tau_e)]p(1)\} \\ &\quad /(1-\tau_e p(1)). \end{aligned}$$

同理可得

$$\begin{aligned} V_B &= p(y_B) \\ &= p(x_B)(1-\tau_c)(1-\tau_e)/(1-\tau_e p(1)). \end{aligned}$$

但是 $x_A = x_B$, 因此 (14) 成立。 □

这一定理把 V_A 和 V_B 之间的大小问题归结为税率之间的关系。当 $(1-\tau_b) > (1-\tau_c)(1-\tau_e)$ 时，即债权所得的个人所得税率相比于股权所得的个人所得税率要低得多时，看来发行债券的公司比不发行债券的公司的价值要高。反之，当这个不等式相反时，结论也相反。而当 $(1-\tau_b) = (1-\tau_c)(1-\tau_e)$ 时，公司的价值又与债权/股权比无关。

沿着类似的思路我们还可作更深入的分析。不过所有这些都是建立在理想的市场环境下。这一市场首先是完全透明的，即对每一投资者来说，都能掌握市场的所有信息。实际情况当然并非如此，尤其是公司的债权/股权比更是内部信息，外人往往无法得知。这样，如果要考虑信息上的不透明性，或者说信息的不对称性，那么上述讨论就更显得很不足。一种简单的考虑就是把债权/股权比作为一个影响现金流估值的因素，于是就有可能通过一种“均衡”分析，来确定最优的债权/股权比。与此有关的研究可参看 Ross [11]。这方面还在继续引起许多更有说服力的研究，使公司财务这个学科不断深入发展。

从上面的介绍中我们可以看到，所谓 Modigliani–Miller 定理，归根到底只是线性定价法则的简单应用。因而，关于它的有关讨论是否符合实际的问题也就是公司价值是否满足线性定价法则的问题。而后一问题无非就是公司的总价值是否就等于它的局部价值的总和。这一问题的答案当然在一定程度上应该是对的，但是相反的情况同样到处可见。甚至人们可以说，用各种渠道融资来创建一个公司的动力就在于要使其总体价值大于其局部价值之和。尤其是公司之间大量的收购并购现象正是说明线性定价法则在这类情况下一定不成立。最近的一篇关于公司财务理论发展的综述 [13] 指出，从理论上来说，Modigliani–Miller 定理是否成立在于怎样来理解一个公司。如果把公司理解为一个“明确的合约的连接 (a nexus of explicit contracts)”，那么 Modigliani–Miller 定理理所当然地应该成立。但是现实说明许多公司都没有那么简单。即使简单地对公司考虑其“隐含合约 (implicit contract)” (例如，债权和股权就有不同的“隐含合约”) 的存在，Modigliani–Miller 定理就已经需要修正。更何况对于现代企业来说，还有许多重要的决定性的因素 (例如人力资本、公司治理结构等)需要考虑。尽管如此，Modigliani–Miller 定理作为公司财务理论的奠基性研究，其重要性仍然是不可动摇的。

说明 Modigliani–Miller 定理 (MMT) 是由 Modigliani 和 Miller 于 1958 年在他们的经典论文 [9] 中明确提出的。在目前流行的文献中，有时也称它们为 Modigliani–Miller 命题 (MM Proposition, MMP)。通常把我们的定理 3.3：“公司价值与资本结构无关”称为 MMT 1 或 MMP 1，而 MMT 2 或 MMP 2 是指“公司的股权资本费用是债权/股权比的线性递增函数”。即定理 3.3 的推论。这两个结果都出现在 [9] 中。至于我们这里的定理 3.1，出现在 Modigliani 和 Miller 在 1961 年发表的论文 [10] 中。它有时被称为 MMT 3。此外，还有习题 6 的叙述有时也称为 MMT 4。有关税收的讨论则主要来自 Miller [8]。

通过这里的介绍，我们应该认识到 MM 定理是在什么假设下得到的。我们这里没有像许多教科书中列出“资本市场高度完善，利率一致”之类的说明性的假定，因为所提出的数学模型中其实已经很明确地包含这些假定。尤其是在早期的文献中，完全市场的条件是通过同一个风险类的公司足够多来建立的。现在我们知道完全市场的条件不必通过公司类来建立。这一点首先由 Stiglitz [12] 意识到。他提出，可用 Arrow–Debreu 证券集存在的“Arrow–Debreu 市场”来取代风险类公司的条件。他甚至在那里已提出资产定价基本定理的雏形。然而，我们也应该时时意识到数学抽象对于实际意味着什么。MM 定理的实际检验是很困难的，因为很难找到两个有同样状况的公司。这就使得它更像是一种一般均衡框架对公司财务的具体化。

MM 定理是公司财务理论的基础。因此，在以公司财务作为主要目标的 Coperand–Weston [1] 中，无论在理论上还是在实证上，对此都作了详尽讨论 (第 13–16 章)。其中尤其是在理论上，它把 MM 定理与 CAPM 和期权定价理论相联系的讨论值得注意。但在一般的金融经济学教科书中，对 MM 定理往往不作重点介绍。 Huang–Litzenberger [4] 中仅在 128 页不显眼地提到它。Ingersoll [5] 则只对它在“连续金融

学”的情形下进行讨论。我们的介绍基本上来自 Jarrow [6] 的有关章节 (第 4，10 和 11 章)。其实，作为线性定价法则的简单应用，MM 定理是非常好、也非常有用的题材。MM 定理在 Duffie [3] 中只出现在 48 页的注解里。在 Duffie [2] 的 122–124 页中，有本质上与这里讨论一致的 MM 定理，但它是在均衡框架中讨论的，并且其表达过分数学化。

思考题和习题

1. 什么是有关分红政策的 Modigliani–Miller 定理？联系我国股市的实际情况，这一定理有什么启示？
2. 什么是 Modigliani–Miller 定理 (定理 3.3)？它是在什么条件下得到的？它的实际意义何在？
3. Modigliani–Miller 定理的证明是怎样紧密联系线性定价法则的？参照有些公司财务的著作，你能给出用一个简单的状态表 (未来只有有限种状态) 来描述 Modigliani–Miller 定理吗？
4. 怎样在 Modigliani–Miller 定理的框架中考虑税收、破产等因素？
5. 给出定理 3.5 的详细证明。
6. 通常还把 MM 定理表达为：“公司股东不关心公司的财务政策”。怎样来理解这一说法？
7. 你能设想在 MM 定理框架中考虑信息的作用吗？

参考文献

[1] Coperand, T. E., and J. F. Weston, 1988, Financial Theory and Corporate Policy, Addison-Wesley.

[2] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[3] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N.J..

[4] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[**5**] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[**6**] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[**7**] Miller, M., 1977, Debt and taxes, *Journal of Finance*, **32**: 261–275.

[**8**] Miller, M. and M. Scholes, 1978, Dividends and taxes, *Journal of Financial Economics*, **6**: 333–364.

[**9**] Modigliani, F., and M. Miller, 1958, The cost of capital, corporation finance, and the theory of investment, *American Economic Review*, **48:** 261–297 (中译文见：莫迪利亚尼文萃， 2001，林少宫、费剑平译，首都经济贸易大学出版社，北京，106–151)。

[**10**] Modigliani, F., and M. Miller, 1961, Dividend policy, growth and the valuation of shares, *Journal of Business*, **34**: 411–433.

[**11**] Ross, S. A., 1977, The determination of financial structure: the incentive-signalling approach, *Bell Journal of Economics*, **32:**, 1467–1484.

[**12**] Stiglitz, J., 1969, A re-examination of the Modigliani–Miller Theorem, *American Economic Review,* **59**:, 784–93.

[**13**] Zingales, L., 2000, In search of new foundations, *Journal of Finance*, **55**: 1623–1653.

四、Markowitz 证券组合选择理论和资本资产定价模型

本讲要求：了解 Markowitz 均值－方差证券组合选择问题的数学表示及其求解方法。深入理解证券组合的有效前沿和二基金分离定理的经济学含义及其与资本资产定价模型之间的关系。深入理解证券组合选择理论、资本资产定价模型以及随机折现因子理论三者之间的等价关系及其意义。了解允许卖空与不允许卖空两种情形对于证券组合选择理论的区别。

数学预备知识：线性代数。初等概率论 (随机变量的数学期望、方差、协方差等)。初等微分学（求导数，条件极值的必要条件)。

内容：

- 4.1 证券组合的收益率和证券组合选择问题。
- 4.2 两种证券的证券组合选择问题。
- 4.3 协方差矩阵正定的一般情形下的均值－方差证券组合选择问题的解。
- 4.4 带无风险证券的均值－方差证券组合选择问题的解。
- 4.5 二基金分离定理与资本资产定价模型。
- 4.6 证券组合选择理论、资本资产定价模型和随机折现因子理论的等价性。
- 4.7 不允许卖空的均值－方差证券组合选择问题。
- 附录 1：资本资产定价模型的 Sharpe 证明
- 附录 2：数学预备知识 2。

正如我们在“代引言”中所述，开始于 1952 年的 Markowitz 的证券组合选择理论是现代金融经济学的起点。其中在观念上最重要的一点是：把证券收益率的方差或标准差作为证券收益风险的度量。Markowitz 注意到，投资者在选择证券组合时，并非只考虑期望收益率尽可能大，同时还考虑收益率方差尽可能小。由此他就提出所谓“期望收益－收益方差 (expected return–variance of return)”法则，并且认为投资者是按照这一法则来进行投资的。

4.1　我们首先来考虑 Markowitz 证券组合选择问题所涉及的数学模型。Markowitz 讨论的证券市场中只有有限种证券，并且他并没有考虑证券的未来价格，而是直接从证券的收益率出发来提出问题。但是为了便于理解，我们不妨先从考虑证券价格的模型出发，并且假定线性定价法则已经成立，来讨论 Markowitz 提出的问题。然后，再舍弃线性定价法则，直接从证券的收益率出发来讨论，并且指出 Markowtiz 问

题的有效前沿的存在性等价于这个证券市场中的线性定价法则成立。揭示后面所说的逻辑关系，显然能人使我们对 Markowitz 理论有更深刻的理解。

假定有当前和未来两个时刻，当前是确定的，但未来是不确定的。假定市场中有 n 种基本证券，其未来价格是 n 个随机变量 $x_1, x_2, \ldots, x_n$. 这 n 种证券的投资组合可用 n 维向量 $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ 来表示，而投资组合的未来价值为

$$y = \theta \cdot x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n. \tag{1}$$

又假定线性定价法则成立，以至存在某个线性定价函数 p (目前它是定义在随机变量集合上的函数)，使得这 n 种证券的当前价格为 $p(x_1), p(x_2)$, ..., $p(x_n)$. 而证券组合 θ 的当前价值则是

$$p(y) = p(\theta \cdot x) = \theta_1 p(x_1) + \theta_2 p(x_2) + \cdots + \theta_n p(x_n).$$

由此可得到该证券组合的 (总) 收益率 r_y 为

$$\begin{aligned} r_y &= \frac{y}{p(y)} \\ &= \frac{\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n}{p(\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n)} \\ &= \frac{\theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n}{\theta_1 p(x_1) + \theta_2 p(x_2) + \cdots + \theta_n p(x_n)} \\ &= w_1 r_1 + w_2 r_2 + \cdots + w_n r_n, \end{aligned} \tag{2}$$

其中

$$r_i = \frac{x_i}{p(x_i)}, \qquad i = 1, 2, \ldots, n;$$

它们是各基本证券的收益率；

$$w_i = \frac{\theta_i p(x_i)}{\theta_1 p(x_1) + \theta_2 p(x_2) + \cdots + \theta_n p(x_n)}, \qquad i = 1, 2, \ldots, n.$$

注意到分母是该证券组合的当前价值，故 w_i 是组合中第 i 种证券的当前价值在其中所占的比例 (不过它有可能是负的)。尤其是

$$w_1 + w_2 + \cdots + w_n = 1. \tag{3}$$

从这一段叙述来看，如果线性定价法则成立，一个证券组合的收益率与基本证券收益率之间的关系是很明确的，即证券组合的收益率等于基本证券收益率的“仿射组合” (系数和等于 1 的线性组合)，并且

“仿射组合”的系数 w_i 就等于第 i 种证券的当前价值在证券组合的当前价值中所占的比例。然而，Markowitz 理论并不需要线性定价法则必须成立。事实上，在 Markowitz 1952 年的经典文献 [9] 中，一切讨论都是直接从证券和证券组合的收益率出发的。从数学公理化的观点来看，我们应该把这里的假设前提进一步明确。对此，我们提出下列定义：

Markowitz 组合收益集 设 $r_1, r_2, \ldots, r_n$ 为 n 个方差有限的随机变量，它们称为 n 种证券的 **收益率**。下列集合 R_1 中的元素称为这 n 种证券的**组合的收益率**：

$$R_1 = \left\{ r = w_1 r_1 + w_2 r_2 + \cdots + w_n r_n \,|\, w_i \in \mathbb{R},\, i = 1, \ldots, n;\ \sum_{i=1}^{n} w_i = 1. \right\}$$

这里我们完全不提证券的当前价格以及定价函数。因此，这个收益率集合 R_1 也不再有“当前价值为 1 的未定权益全体”的意义。我们以后的结论是：如果 Markowitz 问题的有效前沿存在，那么在适当的条件下，一定存在某个线性定价函数，使得这一收益率集合仍然又有我们前面所说的意义。

现在我们要引入证券和证券组合的风险度量。Markowitz 当年的一个重要观念是：风险用收益率的方差或标准差来刻划。如果 $V_{ij} = \mathrm{Cov}[r_i, r_j]$ 是 r_i 和 r_j 之间的协方差，那么 r_y 的标准差 σ_y 应该满足下列公式：

$$\begin{aligned} \sigma_y^2 &= E\left[\left(\sum_{i=1}^{n} w_i r_i - \sum_{i=1}^{n} w_i E[r_i]\right)^2\right] \\ &= \sum_{i,j=1}^{n} w_i w_j E[(r_i - E[r_i])(r_j - E[r_j])] \\ &= \sum_{i,j=1}^{n} V_{ij} w_i w_j. \end{aligned} \tag{4}$$

而 Markowitz 的证券组合选择理论的出发点就是 (2)–(4).

Markowitz 考虑的问题是如何确定 w_i, 使得证券组合在收益率一定时，风险 (收益率的方差或标准差) 最小。为了使表达比较简洁，我们将使用下列矩阵表示：

$$w = (w_1, w_2, \ldots, w_n)^T, \quad e = (1, 1, \ldots, 1)^T,$$

$$\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T, \quad \mu_i = E[r_i], \quad i = 1, 2, \ldots, n,$$

$$V=(V_{ij})_{i,j=1,2,\ldots,n}=(\mathrm{Cov}[r_i,r_j])_{i,j=1,2,\ldots,n},$$

并不妨称 w 为**组合**，$\mu_w=w^T\mu$ 为**组合的收益**，$\sigma_w=(w^TVw)^{1/2}$ 为**组合的风险**。这样，Markowitz 的问题 (它常称为**均值－方差证券组合选择问题**) 为

$$\begin{cases} \min & \sigma_w^2=w^TVw=\sum_{i,j=1}^n V_{ij}w_iw_j \\ \text{s.t.} & w^Te=w_1+w_2+\cdots+w_n=1 \\ & \mu_w=w^T\mu=w_1\mu_1+w_2\mu_2+\cdots+w_n\mu_n=\bar{\mu} \end{cases} \tag{5}$$

这里 min 表示对后面的证券组合的收益率方差 σ_w^2 求最小值，s.t. (subject to 的缩写) 表示约束条件。这一问题的解 $\bar{w}$ 称为对应收益 μ_w 的**极小风险组合**。

用数学的语言来说，这是个二次规划问题，即它是在两个线性等式约束条件下的二次函数的求最小值的问题。作为 n 个随机变量的协方差矩阵 V, 它一定是非负定的，即对于任何 n 维向量 w, 它必然有 $\sigma_w^2=w^TVw\geqslant 0$. 这样，我们面临的最小化函数是 n 个变量的非负定二次函数。两个约束条件又使 n 维变量被限制在一个闭集上。考虑到非负定二次函数的特殊性，可以指出，对于任何固定的 $\bar{\mu}$, 这个问题的解是一定存在的。

如果除了 $w=0$ 外，上面的不等式中的严格不等号始终成立，那么 V 就被称为是**正定的**。正定矩阵一定是非异矩阵，也就是说，它所对应的行列式一定不等于零 (并且一定大于零)，以至它的逆矩阵存在。但非负定的矩阵可能是奇异矩阵，即其对应的行列式可能等于零。我们以后经常会假定 V 是正定的。有关问题我们放到以后去讨论。这里我们先提一下正定矩阵的一个性质：正定矩阵的主子行列式都大于零，尤其是正定矩阵的主对角线上的元素都大于零。这里主子行列式是指包含主对角线元素的行列式。反之，如果一个矩阵有这样的性质，它一定是正定矩阵。这一性质是所谓 Sylvester 定理。实际上，它是由正定矩阵的主子矩阵一定也是正定矩阵这一性质导出的。

4.2　现在我们从几何上来想象这一问题的求解。一维的二次正定函数的图象是开口向上的抛物线；高维的二次正定函数的图象是开口向上的抛物超曲面，而两个约束条件表示变量落在两个超平面的交集上，抛物超曲面被这个交集截出一个截口。其最低点就对应问题的解。Markowitz 最初对于 $n=3$ 的情形就是利用这样的几何方法来讨论求解的。

对于 $n=2$, 问题较为简单。在 $\mu_1\neq\mu_2$ 时，由两个约束等式就可解出 w_1 和 w_2 来。事实上，由

$$w_1+w_2=1,\qquad w_1\mu_1+w_2\mu_2=\mu_w,$$

立即可得

$$w_1 = \frac{\mu_w - \mu_2}{\mu_1 - \mu_2}, \quad w_2 = \frac{\mu_w - \mu_1}{\mu_2 - \mu_1}.$$

$$\sigma_w^2 = V_{11}\left(\frac{\mu_w - \mu_2}{\mu_1 - \mu_2}\right)^2 - 2V_{12}\frac{(\mu_w - \mu_1)(\mu_w - \mu_2)}{(\mu_1 - \mu_2)^2} + V_{22}\left(\frac{\mu_w - \mu_1}{\mu_1 - \mu_2}\right)^2.$$

把上式稍加整理可得

$$(\mu_1 - \mu_2)^2\sigma_w^2 = \frac{(V_{11}(\mu_w - \mu_2) - V_{12}(\mu_w - \mu_1))^2 + (V_{11}V_{22} - V_{12}^2)(\mu_w - \mu_1)^2}{V_{11}}.$$

我们假定 V 是正定矩阵，故 $V_{11} > 0$ (它意味着第一种证券收益率的方差不为零), $V_{11}V_{22} - V_{12}^2 > 0$ (它意味着两种证券收益率的相关系数不等于 ±1). 因此，上式右端恒大于零；从而它一定可表示为 $a(\mu_w - b)^2 + c$ 的形式，并且 $a > 0$, $c > 0$. 这样一来，再令 $d = (\mu_1 - \mu_2)^2$, 我们就得到 σ_w 与 μ_w 满足如下形式的关系式：

$$d\sigma_w^2 - a(\mu_w - b)^2 = c, \quad a, c, d > 0.$$

这在 (σ_w, μ_w) 平面上是一条双曲线的图象，并且由于 $\sigma_w \geqslant 0$, 它只有开口向右的半支。

我们还可对上述 $n = 2$ 的情形的特殊情况作出更具体的计算。记 $V_{11} = \sigma_1^2$, $V_{22} = \sigma_2^2$, $V_{12} = \sigma_1\sigma_2 r_{12}$, 其中 σ_1 为第一种证券的收益率的标准差，σ_2 为第二种证券的收益率的标准差，而 r_{12} 为两种收益率的相关系数 (协方差与两个标准差乘积之比)。r_{12} 可能取 -1 和 $+1$ 之间的任何值。如果 $r_{12} = 1$, 即 r_1 和 r_2 完全正相关 (这时 V 不正定)，那么

$$\sigma_w^2 = (\sigma_1 w_1 + \sigma_2 w_2)^2 = \frac{(\sigma_1(\mu_w - \mu_2) - \sigma_2(\mu_w - \mu_1))^2}{(\mu_1 - \mu_2)^2}.$$

对两端开方，并假设 $\mu_1 > \mu_2$, 则得

$$\sigma_w = \frac{\sigma_1(\mu_w - \mu_2) - \sigma_2(\mu_w - \mu_1)}{\mu_1 - \mu_2}.$$

这里为保证 $\mu_w = \mu_1$ 时，$\sigma_w = \sigma_1$ 以及 $\mu_w = \mu_2$ 时，$\sigma_w = \sigma_2$, 正负号只有一种取法。这一方程说明，在 $r_{12} = 1$ 时，上述的双曲线退化为在 (σ_w, μ_w) 平面上的连接点 (σ_1, μ_1) 和 (σ_2, μ_2) 的直线，不过要除去 $\sigma_w < 0$ 的部分。该直线于 μ_w 轴交于点 $(0, (\sigma_1\mu_2 - \sigma_2\mu_1)/(\sigma_1 - \sigma_2))$. 其中后一个坐标就是在这种情况的无风险证券收益率。如果 $r_{12} = -1$, 即 r_1 和 r_2 完全负相关 (这时 V 也不正定)，那么

$$\sigma_w^2 = (\sigma_1 w_1 - \sigma_2 w_2)^2 = \frac{(\sigma_1(\mu_w - \mu_2) + \sigma_2(\mu_w - \mu_1))^2}{(\mu_1 - \mu_2)^2}.$$

对两端开方，并假设 $\mu_1 > \mu_2$, 那么为保证 $\mu_w = \mu_1$ 时，$\sigma_w = \sigma_1$ 以及 $\mu_w = \mu_2$ 时，$\sigma_w = \sigma_2$, 正负号的两种取法都有可能。其中通过点 (σ_1, μ_1) 的直线为

$$\sigma_w = \frac{\sigma_1(\mu_w - \mu_2) + \sigma_2(\mu_w - \mu_1)}{\mu_1 - \mu_2}.$$

而通过点 (σ_2, μ_2) 的直线为

$$\sigma_w = \frac{\sigma_2(\mu_w - \mu_1) + \sigma_1(\mu_w - \mu_2)}{\mu_2 - \mu_1}.$$

这两条直线交于 μ_w 轴的同一点 $(0, (\sigma_1\mu_2 + \sigma_2\mu_1)/(\sigma_1 + \sigma_2))$, 其中后一个坐标是在这种情况下的无风险利率。这两种情形的讨论使我们很容易构造出不满足线性定价法则的三种证券的收益率来，即通过它们的组合可以得到两种不同的无风险利率。这说明，任意的一组随机变量不一定是一些证券的收益率，它们有可能与线性定价法则矛盾。另一方面，这两种完全正负相关情形都使 V 不正定。要求 V 正定的原因之一也正是为了回避这两种情况。

还有一种特殊情形是 $r_{12} = 0$, 即 r_1 与 r_2 不相关。这时，

$$\sigma_w^2 = (\sigma_1 w_1)^2 + (\sigma_2 w_2)^2 = \frac{\sigma_2^2(\mu_w - \mu_1)^2 + \sigma_1^2(\mu_w - \mu_2)^2}{(\mu_1 - \mu_2)^2}.$$

它在 (σ_w, μ_w) 平面上是通过点 (σ_1, μ_1) 和 (σ_2, μ_2) 的双曲线。我们可以求出 σ_w 达到最小值的点 (即双曲线的顶点) 为$(\sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2), (\mu_1\sigma_2^2 + \mu_2\sigma_1^2)/(\sigma_1^2 + \sigma_2^2))$. 这里可以注意到，只要 σ_1 和 σ_2 都不为零，我们有

$$\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} < \min\{\sigma_1^2, \sigma_2^2\}.$$

这就是说，对于任何两种收益率不相关的证券，总能给出一种风险比原来两者都要小、而收益率在两者之间的证券组合来。这里没有提到的其他中间情形可以类推。它们所对应的协方差矩阵 V 都是正定的，而对应的图象都是双曲线。

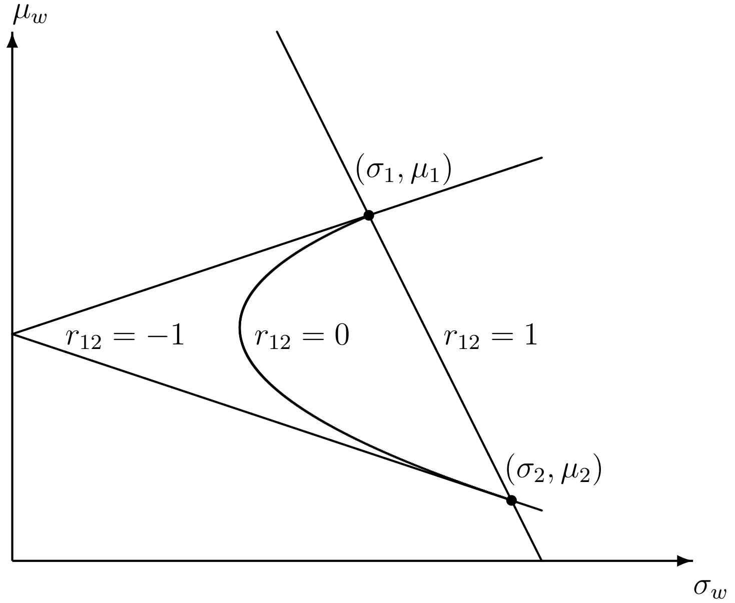


图 1: 两种证券的组合

4.3 现在我们来讨论一般情形。为了使问题便于讨论，我们**假设 V 正定**，即对于任何 $w \neq 0$, 有 $w^T V w > 0$. 正定矩阵一定是非异的，即其对应的行列式不为零，以致其逆矩阵 V^{-1} 存在。在这样的假设下，首先排除了这 n 种证券中有无风险证券，否则它的收益率将与其他证券的收益率都不相关，会使 V 出现全为零的行与列。其次是不存在非零组合 w, 使得其对应的组合风险 σ_w 为零。这意味着不存在某些证券的组合等同于一种无风险证券；或者说，这些风险证券的任何组合仍然是风险证券。尤其是这样将排除把两种收益率完全相反的证券放在一起来考虑。最后，还排除了两种证券的收益率完全相关的情形，例如，不允许有 $\mathrm{Cov}^2[r_1, r_2] = \mathrm{Var}[r_1]\mathrm{Var}[r_2]$, 即不允许 $V_{11}V_{22} - V_{12}^2 = 0$, 因为 V 作为正定矩阵其主子行列式都应大于零。不过 V 正定这一条件其实还是为了数学讨论上的方便，并不意味着现实中不可能出现 V 非正定的情况，也不意味着当 V 非正定时问题就没有解[1]。同时，我们还要求所有证券的**期望收益率 μ_i 不能都相等**。否则 μ_w 只能恒等于这公共的期望收益率。这是比较容易讨论的情形。

如前面所述，Markowitz 问题 (5) 是一个二次规划问题。这一问题用普通的 Lagrange 乘子法就能求得其解。(5) 的 Lagrange 函数为

$$L(w, \lambda_1, \lambda_2) = w^T V w - \lambda_1(w^T \mu - \bar{\mu}) - \lambda_2(w^T e - 1).$$

---

[1] V 不正定只是说明所考虑的证券中有些是“多余的”。拿走“多余的”证券，余下的证券仍能生成原来能生成的各种组合收益率。而余下的证券集的 V 将正定。

其解 $w=\bar{w}$ 满足的必要条件为

$$\frac{\partial L(\bar{w},\lambda_1,\lambda_2)}{\partial w}=2Vw-\lambda_1\mu-\lambda_2 e=0, \tag{6}$$

$$\frac{\partial L(\bar{w},\lambda_1,\lambda_2)}{\partial \lambda_1}=\bar{\mu}-\bar{w}^T\mu=0, \tag{7}$$

$$\frac{\partial L(\bar{w},\lambda_1,\lambda_2)}{\partial \lambda_2}=1-\bar{w}^T e=0, \tag{8}$$

这里 (6) 是个 n 维向量等式。由于这是个二次凸规划问题，还可以指出，这一方程组也是 $\bar{w}$ 为 (5) 的解以及 λ_1,λ_2 为 (5) 的 Lagrange 乘子的充分条件 (参看附录 2)。

由此可解得

$$\bar{w}=\frac{1}{2}V^{-1}(\lambda_1\mu+\lambda_2 e)=\frac{1}{2}V^{-1}(\mu\ e)\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}.$$

这里我们把中间的表达式记成矩阵相乘的形式。另一方面，又有

$$(\mu\ e)^T\bar{w}=\begin{pmatrix}\bar{\mu}\\ 1\end{pmatrix},$$

从而

$$(\mu\ e)^T\bar{w}=\frac{1}{2}(\mu\ e)^TV^{-1}(\mu\ e)\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}=\begin{pmatrix}\bar{\mu}\\ 1\end{pmatrix}.$$

令

$$A=(\mu\ e)^TV^{-1}(\mu\ e).$$

如果这个二阶矩阵的逆 A^{-1} 存在，那么由此可得

$$\frac{1}{2}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}=A^{-1}\begin{pmatrix}\bar{\mu}\\ 1\end{pmatrix},$$

以至

$$\bar{w}=\frac{1}{2}V^{-1}(\mu\ e)\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}=V^{-1}(\mu\ e)A^{-1}\begin{pmatrix}\bar{\mu}\\ 1\end{pmatrix}. \tag{9}$$

余下的是还要指出 A^{-1} 的存在性。对此我们只需指出 A 是正定矩阵。事实上，对于任何不同时为零的实数 y_1,y_2, 有

$$\begin{aligned}(y_1\ y_2)A\begin{pmatrix}y_1\\ y_2\end{pmatrix} &= (y_1\ y_2)(\mu\ e)^TV^{-1}(\mu\ e)\begin{pmatrix}y_1\\ y_2\end{pmatrix}\\ &= (y_1\mu+y_2e)^TV^{-1}(y_1\mu+y_2e)>0,\end{aligned}$$

这是因为 V^{-1} 是正定矩阵。由此导得 A 是正定的。

最后，我们把上述讨论结果写成定理形式。

定理 4.1　在均值－方差证券组合选择问题 (5) 中，如果 n 种证券的期望收益率 μ_i 不全相同，收益率协方差矩阵 V 正定，组合的期望收益率 $\bar{\mu}$ 给定，那么 (5) 有唯一解 $\bar{w}$ 为

$$\bar{w} = V^{-1}(\mu\ e)A^{-1}\begin{pmatrix} \bar{\mu} \\ 1 \end{pmatrix}. \tag{10}$$

其中

$$A = (\mu\ e)^T V^{-1}(\mu\ e). \tag{11}$$

令

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

那么对于组合的期望收益率 $\bar{\mu}$ 和最小方差 $\bar{\sigma}^2$ 之间应该满足下列表达式：

$$\begin{aligned}
\bar{\sigma}^2 &= \bar{w}^T V \bar{w} = (\bar{\mu}\ 1)A^{-1}(\mu\ e)^T V^{-1} V V^{-1}(\mu\ e)A^{-1}\begin{pmatrix} \bar{\mu} \\ 1 \end{pmatrix} \\
&= (\bar{\mu}\ 1)A^{-1}\begin{pmatrix} \bar{\mu} \\ 1 \end{pmatrix} \\
&= (\bar{\mu}\ 1)\frac{1}{(ac-b^2)}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}\begin{pmatrix} \bar{\mu} \\ 1 \end{pmatrix} \\
&= \frac{a - 2b\bar{\mu} + c\bar{\mu}^2}{(ac-b^2)}.
\end{aligned}$$

这说明在 (σ_w, μ_w) 平面上，极小风险组合 $\bar{w}$ 的收益 $\bar{\mu}$ 与风险 $\bar{\sigma}$ 之间画出了一条双曲线的右半支 (如图 2)，它称为组合的**前沿**。其中双曲线的顶点对应**总体最小方差组合** w_G. 容易直接由上式计算得到其对应的组合收益 $\mu_G = b/c$, 组合风险为 $\sigma_G = 1/\sqrt{c}$, 而

$$\begin{aligned}
w_G &= V^{-1}(\mu\ e)A^{-1}\begin{pmatrix} \mu_G \\ 1 \end{pmatrix} \\
&= \frac{1}{(ac-b^2)}V^{-1}(\mu\ e)\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}\begin{pmatrix} b/c \\ 1 \end{pmatrix} = \frac{V^{-1}e}{c}.
\end{aligned}$$

双曲线的上半部称为**有效前沿**，有效前沿上的每一点所对应的组合称为**有效组合**。它们都代表一种收益固定时风险最小的组合以及风险固定时收益最大的组合。把收益和风险看成组合两个目标，那么有效组

合对应所谓 Pareto 最优组合，即对它们来说，不可能找到两方面都比它好的组合。而在它们之间互相比较，则可发现**高收益对应高风险**。双曲线的下半部称为**无效前沿**，在无效前沿上的每一点也都代表一种收益固定时风险最小的组合，但它们的收益在风险固定时并不是最大。对于投资者来说，当然他只关心有效前沿。这对他的决策十分重要。他将根据他自己的投资偏好在有效前沿上选取他的投资组合。

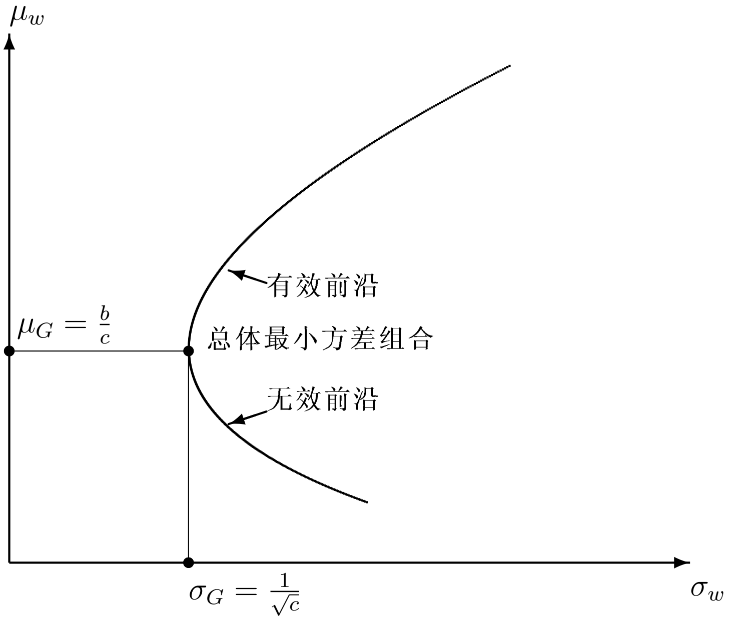


图 2: n 种证券的组合

在 Markowitz 提出他的理论时，要计算有意义的证券组合的有效前沿，那只能纸上谈兵。甚至到了 1980 年代，计算上千种证券的组合有效前沿，也还相当困难。因此，计算有效前沿的问题曾经是 Sharpe 提出资本资产定价模型的一个动机[2]。但是随着计算技术的发展，目前要计算几千种证券的组合有效前沿都可在几分钟内完成。

4.4 以上的讨论中由于要求 V 正定，使得 n 种证券中不可能包括无风险证券。对于有无风险证券情形，我们可以假设除上述 n 种证券外，另外还有第 0 种证券为无风险证券，并且它的无风险利率为常随机变量 r_f. 在这种情况下，组合将定义为满足 $w_0 + w_1 + \cdots + w_n = 1$ 的

---

[2] 如果 $r_i - r_f = \beta_i(r_m - r_f) + \varepsilon_i$, $i = 1, 2, \ldots, n$, 其中 ε_i 与 r_m 不相关，且 ε_i 之间互不相关，那么 $\mathrm{Cov}[r_i, r_j] = \beta_i\beta_j\mathrm{Var}[r_m]$; 从而协方差矩阵的计算量从 n^2 的数量级降到 n 的数量级。

$w_0, w_1, \ldots, w_n$. 我们仍记 $w = (w_1, \ldots, w_n)^T$. 这时，组合的期望收益率为

$$\mu_w = w_0 r_f + w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n,$$

从而

$$\mu_w - r_f = w_1(\mu_1 - r_f) + w_2(\mu_2 - r_f) + \cdots + w_n(\mu_n - r_f) = w^T(\mu - r_f),$$

而组合的方差则显然仍为 $\sigma_w^2 = w^T V w$. 令 $\mu_i' = \mu_i - r_f$, $\mu_w' = \mu_w - r_f$. 那么现在的均值－方差组合选择问题变为

$$\begin{cases} \min & \sigma_w^2 = w^T V w = \sum_{i,j=1}^n V_{ij} w_i w_j \\ \text{s.t.} & \mu_w' = w^T \mu' = w_1\mu_1' + w_2\mu_2' + \cdots + w_n\mu_n' = \bar{\mu}' \end{cases} \tag{12}$$

也就是说，形式上与原来的问题的区别仅在于少了一个约束条件。(12) 的 Lagrange 函数为

$$L(w, \lambda) = w^T V w - \lambda(w^T\mu' - \bar{\mu}').$$

其解 $w = \bar{w}$ 满足的必要条件为

$$\frac{\partial L(\bar{w}, \lambda)}{\partial w} = 2V\bar{w} - \lambda\mu' = 0, \tag{13}$$

$$\frac{\partial L(\bar{w}, \lambda)}{\partial \lambda} = \bar{\mu}' - \bar{w}^T\mu' = 0, \tag{14}$$

由此可解得

$$\bar{w} = \left(\frac{\bar{\mu}'}{(\mu')^T V^{-1}\mu'}\right) V^{-1}\mu'. \tag{15}$$

而相应的极小方差为

$$\begin{aligned} \bar{\sigma}^2 &= \bar{w}^T V \bar{w} \\ &= \left(\frac{\bar{\mu}'}{(\mu')^T V^{-1}\mu'}\right)^2 (\mu')^T V^{-1} V V^{-1}\mu' \\ &= \frac{(\bar{\mu}')^2}{(\mu')^T V^{-1}\mu'}. \end{aligned}$$

这就是说，$\bar{\sigma}$ 与 $\bar{\mu}' = \bar{\mu} - r_f$ 之间在 (σ_w, μ_w) 平面上的双曲线关系在这种情形退化为两条直线：

$$\bar{\sigma} = \frac{\pm\bar{\mu}'}{((\mu')^T V^{-1}\mu')^{1/2}}.$$

由于 $\bar{\sigma}$ 必须为正，这两条直线也都只有右边的半条射线，并相交于 μ_w 轴上的 r_f 点 (如图 3)。其中上半条射线是组合的有效前沿；下半条射线是组合的无效前沿。从经济意义上来看，无风险利率 r_f 与总体最小风险组合的期望收益率 μ_G 相比，应该要小。因此，它们之间的位置应该如图 3 中所示。但从理论上来说，似乎没有理由说 $r_f \geqslant \mu_G$ 一定不可能。由 (15)，如果组合收益率 $\bar{\mu}$ 满足

$$\bar{\mu}' = \frac{\mu'^T V^{-1} \mu'}{e^T V^{-1} \mu'},$$

那么 $\bar{w}$ 满足 $\bar{w}^T e = 1$，即这时组合中无风险证券的比例为零。因此，它必须对应 n 种风险证券组合的前沿上的一点，即相应的双曲线必须与射线相切于该组合对应的点。当 $e^T V^{-1} \mu' > 0$ 时，该切点 (σ_m, μ_m) 在上半条射线上 (如图 3)；这时 $r_f < \mu_G$. 当 $e^T V^{-1} \mu' < 0$ 时，该切点在下半条射线上；这时 $r_f > \mu_G$. 当 $e^T V^{-1} \mu' = 0$ 时，实际上说明这样的 $\bar{w}$ 不存在；这时 $r_f = \mu_G$, 而两条射线变为双曲线的渐近线。

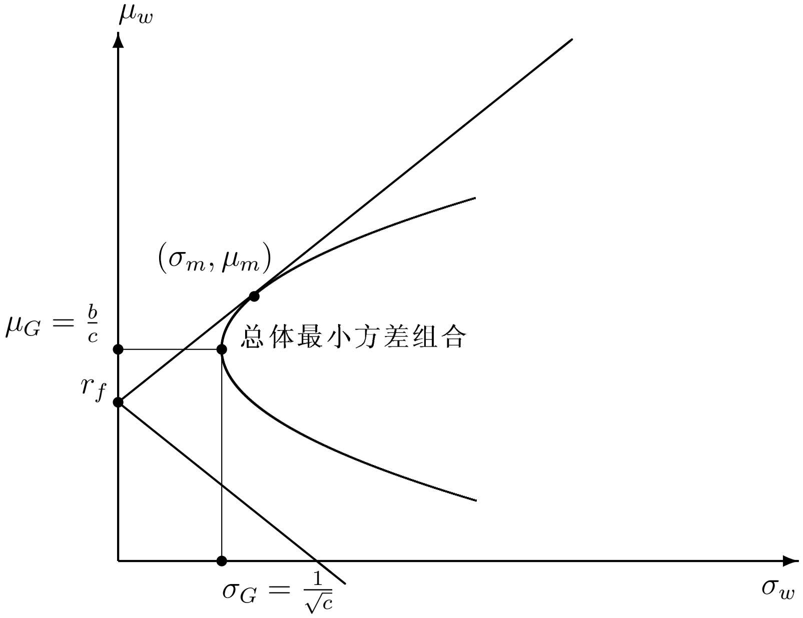


图 3: 无风险证券与 n 种证券的组合

如上所述，带无风险证券的证券组合有效前沿是一条射线。这条直线常称为**资本市场线** (Capital Market Line, CML)。射线的斜率为 $(\mu'^T V^{-1} \mu')^{1/2}$. 这意味着有效前沿上的每一组合 $\bar{w}$ 的期望收益率 $\mu_{\bar{w}}$ 和标准差 $\sigma_{\bar{w}}$ 之间有如下关系：

$$\frac{\mu_{\bar{w}} - r_f}{\sigma_{\bar{w}}} = (\mu'^T V^{-1} \mu')^{1/2}.$$

左端的比值称为组合的 **Sharpe 比**, 它用来衡量**风险效益**，即因承担风险而可能带来的收益。带无风险证券的证券组合有效前沿的特点就在于其上的 Sharpe 比是常数 $(\mu'^T V^{-1}\mu')^{1/2}$, 它完全由各风险证券的期望收益率 μ 和它们间的协方差矩阵 V 所决定。一般情况下，有效前沿射线与余下的风险证券组合的有效前沿相切于一点 (σ_m, μ_m). 因此，在这条射线上的每一点所对应的期望收益 $\mu_{\bar{w}}$ 都有下列形式：

$$\mu_{\bar{w}} - r_f = \beta_{\bar{w}}(\mu_m - r_f), \tag{16}$$

其中 $\beta_{\bar{w}} = \sigma_{\bar{w}}/\sigma_m$. 这一等式也有资本资产定价模型的形式，但它并不是资本资产定价模型，而只是说，对应各种有正 β 的证券组合总存在有同样收益的有效前沿上的组合。如果把 (16) 理解为 $\mu_{\bar{w}}$ 与 $\beta_{\bar{w}}$ 之间的关系，那么它的图象也是一条直线。这条直线通常称为**证券市场线** (Security Market Line, SML)。

4.5　我们已经导得 Markowitz 均值－方差证券组合选择问题 (5) 的解。由于这一问题的解答可归结为线性方程组的求解，因此，解的集合有一个特殊的性质。具体地说，就是所得到的极小风险组合满足“叠加原理”，即极小风险组合的仿射组合仍然是极小风险组合。这里所谓**仿射组合**是这样定义的：如果向量 y 与向量 $x_1, x_2 \ldots, x_n$ 间满足下列关系： $y = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n$, $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$, 那么 y 称为 x_1, $x_2, \ldots, x_n$ 的仿射组合。如果这里所有 $\lambda_i \geqslant 0$, 那么 y 称为 $x_1, x_2, \ldots, x_n$ 的**凸组合**。把这一“叠加原理”重新表达，其完整的表述就是下列重要的**二基金分离定理**：

定理 4.2　设组合 w_p 和 w_q 分别是均值－方差组合选择问题 (5) (或 (12)) 的对于期望收益率 $\bar{\mu}$ 为 μ_p 和 μ_q 的解，并且 $\mu_p \neq \mu_q$. 同时，假定定理 4.1 (或带无风险证券情形) 中的条件成立。那么 $\bar{w}$ 是极小风险组合的充分必要条件为存在实数 λ, 使得 $\bar{w} = (1-\lambda)w_p + \lambda w_q$. 如果 w_p 和 w_q 都是有效组合，而 λ 在 0 和 1 之间，那么 $\bar{w} = (1-\lambda)w_p + \lambda w_q$ 也是有效组合。

证明　令 $\bar{\mu} = (1-\lambda)\mu_p + \lambda\mu_q$. 那么利用解的表达式 (10) (或 (15)), 容易证明，$\bar{w} = (1-\lambda)w_p + \lambda w_q$ 就是问题 (5) (或 (12)) 的解。于是问题就简单地归结为由 λ 确定 $\bar{\mu}$ 和由 $\bar{\mu}$ 确定 λ. 至于定理的后一部分可由 λ 在 0 和 1 之间得到 $\bar{\mu}$ 在 μ_p 和 μ_q 之间，从而当 w_p 和 w_q 都对应有效前沿上的组合时，$\bar{w}$ 也将对应有效前沿上的组合。 □

推论　设组合 $w^1, w^2, \ldots, w^k$ 分别是均值－方差组合选择问题 (5) (或 (12)) 的对于期望收益率 $\bar{\mu}$ 为 $\mu^1, \mu^2, \ldots, \mu^k$ 的解，并且 μ^i 之间不全相等。同时，假定定理 4.1 (或带无风险证券情形) 中的条件成立。那么 w^1, $w^2, \ldots, w^k$ 的仿射组合 $\bar{w} = \lambda_1 w^1 + \lambda_2 w^2 + \cdots + \lambda_k w^k$, $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$ 也

是极小风险组合。如果 $w^1, w^2, \ldots, w^k$ 都是有效组合，而 $\lambda_1, \lambda_2 \ldots, \lambda_k \geqslant 0$, 那么 $w^1, w^2, \ldots, w^k$ 的凸组合 $\bar{w} = \lambda_1 w^1 + \lambda_2 w^2 + \cdots + \lambda_k w^k$ 也是有效组合。

证明可以用常规的数学归纳法来进行。

定理 4.2 的含义在于：如果投资者的证券投资决策就是要根据他本人的财力和风险承受能力在均值－方差问题的最优解中选取一点，那么他考虑全体证券的组合与考虑证券的两种组合的组合是一样的。这两种组合在现实证券市场中可能就是两种业绩良好的共同基金 (mutual fund)，因此，这就是说，投资者不必考虑全体证券如何组合，只需考虑如何搭配这两种基金的组合即可。这也是这一定理为什么称为“二基金分离定理”的原因。1981 年诺贝尔经济奖获得者 J. Tobin (1918–) 在 1958 年发表的 [14] 中正是利用一种无风险资产与一种特殊的风险资产的组合来论证类似的二基金分离定理成立，从而得到一些宏观经济方面的结论。为此，人们有时认为 Tobin 也是证券组合分析理论的创始人。

二基金分离定理 4.2 告诉我们的是：由两个极小风险组合的组合就可生成 n 种证券的整个组合前沿。尤其是如果这两个组合也看成是两种证券，那么有了这两种证券以后，为生成同样的组合前沿，其他证券就变得“多余”。把这一想法表达成数学形式，我们得到如下定理。尤其是我们由此再次导得资本资产定价模型。

定理 4.3 设 p 和 q 是两种证券，并且它们的期望收益率 $\mu_p \neq \mu_q$. 那么任何证券 i 不改变 p 和 q 所生成的组合前沿的充分必要条件为：存在实数 $\alpha \in \mathbb{R}$, 使得

$$\mu_i = (1-\alpha)\mu_p + \alpha\mu_q, \tag{17}$$

$$\mathrm{Cov}[r_i, r_p] = (1-\alpha)\mathrm{Var}[r_p] + \alpha \mathrm{Cov}[r_p, r_q], \tag{18}$$

$$\mathrm{Cov}[r_i, r_q] = (1-\alpha)\mathrm{Cov}[r_p, r_q] + \alpha \mathrm{Var}[r_q]. \tag{19}$$

证明 **必要性** 不妨假设 $p = 1$, $q = 2$, $i = 3$. 如果第 3 种证券不改变前两种证券的组合所生成的组合前沿，那么对于任何 $\bar{\mu} \in \mathbb{R}$, 方程 (6)–(8) 对于 $n = 3$ 有 $w_3 = 0$ 的解，即存在实数 $\bar{w}_1, \bar{w}_2$, 和 λ_1, λ_2, 满足

$$\begin{aligned} \bar{w}_1 + \bar{w}_2 &= 1 \\ \mu_1\bar{w}_1 + \mu_2\bar{w}_2 &= \bar{\mu} \\ V_{11}\bar{w}_1 + V_{12}\bar{w}_2 - \frac{1}{2}\lambda_1\mu_1 - \frac{1}{2}\lambda_2 &= 0 \end{aligned}$$

$$
\begin{aligned}
V_{21}\bar{w}_1 + V_{22}\bar{w}_2 - \frac{1}{2}\lambda_1\mu_2 - \frac{1}{2}\lambda_2 &= 0 \\
V_{31}\bar{w}_1 + V_{32}\bar{w}_2 - \frac{1}{2}\lambda_1\mu_3 - \frac{1}{2}\lambda_2 &= 0
\end{aligned}
$$

但是由于 $\mu_1 \neq \mu_2$, 前四个方程对 $\bar{w}_1, \bar{w}_2$, 和 λ_1, λ_2 有唯一解，这说明前四个方程的解一定是第五个方程的解。同时，由于 $\bar{\mu}$ 可取任何实数，我们把第二个方程去掉，仍然可以说，如果 $\bar{w}_1, \bar{w}_2$, 和 λ_1, λ_2 满足第一、三、四个方程，那么它一定满足第五个方程。这仅当第五个方程是第一、三、四个方程的线性组合时才有可能，即存在 $\alpha_1, \alpha_2, \alpha_3$, 使得

$$
\begin{aligned}
\alpha_1 + \alpha_2 V_{11} + \alpha_3 V_{12} &= V_{31} \\
\alpha_1 + \alpha_2 V_{21} + \alpha_3 V_{22} &= V_{32} \\
\alpha_2\mu_1 + \alpha_3\mu_2 &= \mu_3 \\
\alpha_2 + \alpha_3 &= 1 \\
\alpha_1 &= 0
\end{aligned}
$$

令 $\alpha_3 = \alpha$, 则 $\alpha_2 = (1-\alpha)$. 注意到 p, q, i 和 V_{ij} 的定义，立即可得 (17)–(19).

充分性　定理条件实际上指出，上面的方程组中，如果 $\bar{w}_1, \bar{w}_2$, 和 λ_1, λ_2 满足第一、三、四个方程，那么它一定满足第五个方程。考虑到方程 (6)–(8) 的解一定是 Markowitz 问题的解，这说明由前两种证券的组合生成的组合前沿一定也是三种证券的组合生成的组合前沿。这就是说第三种证券不改变前两种证券的组合前沿。 □

推论 (资本资产定价模型)　设证券 p 和 q 满足定理 4.3 的假设，并且 $\mathrm{Cov}[r_p, r_q] = 0$. 那么任何证券 i 不改变 p 和 q 所生成的组合前沿的充分必要条件为其收益率 r_i 满足下列“一般资本资产定价模型”:

$$
E[r_i] - E[r_p] = \beta_i^p(E[r_q] - E[r_p]), \qquad \beta_i^p = \mathrm{Cov}[r_i, r_q]/\mathrm{Var}[r_q]. \tag{20}
$$

特别是当证券 p 为“市场组合” m 时，并把 q 记作 x, (20) 变为**零 β 资本资产定价模型**

$$
E[r_i] - E[r_x] = \beta_i(E[r_m] - E[r_x]), \qquad \beta_i = \mathrm{Cov}[r_i, r_m]/\mathrm{Var}[r_m]. \tag{21}
$$

当证券 x 是无风险证券时，(20) 变为通常的**资本资产定价模型**

$$
E[r_i] - r_f = \beta_i(E[r_m] - r_f), \qquad \beta_i = \mathrm{Cov}[r_i, r_m]/\mathrm{Var}[r_m]. \tag{22}
$$

其中 r_f 是无风险利率。

这一定理及其推论从另一个角度提出资本资产定价模型。这一角度是这样来看的：假定市场中有若干证券。那么由这些证券的组合所形成的组合前沿代表了这一市场的总体“风险－收益”状况。由二基金分离定理 4.2, 这一市场的总体“风险－收益”状况完全可以由两个前沿组合来决定。而任何证券或证券组合不改变市场总体“风险－收益”状况的充分必要条件为这一证券的收益对这两种有效组合满足“广义资本资产定价模型”(定理 4.3)。

原来的资本资产定价模型是对**市场组合**而言的。所谓市场组合是指市场上所有投资者的证券投资的总和所形成的组合。如果它是有效组合，那么上述定理推论中的 m 就适用于这一市场组合。但是市场组合是否是有效的？对此，Sharpe 认为，如果假设所有投资者都是“理性投资者”，并且他们的投资决策都是按照“均值－方差”的原则来进行的，那么每个投资者的证券选择都形成一个有效组合。而两个有效组合的证券合在一起，一定也形成一个有效组合。这是因为它刚好就形成这两个有效组合的凸组合 (习题 1)。由此也可以导得有限个投资者的所有证券合在一起所形成的证券组合也是有效的；尤其是市场组合是有效的。Sharpe 在他的 1964 年经典论文 [12] 中把这一假设称为“**投资者期望均匀性** (homogeneity of investor expectations)”假设。如果这个假设成立，上述资本资产定价模型就成立。

不但如此，如果市场中没有无风险证券，那么选择与市场组合不相关的有效组合，我们就有相应的“零 β 资本资产定价模型”成立。如果市场中有无风险证券，那么再假定市场组合中不包含无风险证券，就可得到市场组合所代表的点就是资本市场线与风险证券有效前沿的切点 (σ_m, μ_m). 在这种情形下，每一有效组合无非就是无风险证券与市场组合的组合，即每一投资者所选取的有效组合与市场组合的差别仅仅在于无风险证券部分。而这时，通常的资本资产定价模型成立。市场组合中不包含无风险证券这一假定意味着无风险证券是投资者之间的一种借贷行为的合约。因此，所有无风险证券的总和为零，或者说，市场中的所有无风险证券必须结清。

这样我们就用另一种方法导得了资本资产定价模型。这一方法基本上是 Sharpe 等原来的思想。但是 Sharpe 原来的证明是从几何图象上的考虑来得到的 (见附录 1)。在这一推导中，除了 Markowitz 理论需要的理论假设以外，最重要的假设是市场组合的有效性。为使市场组合有效，又需要假设所有投资者的证券投资组合有效。于是这里又蕴含着所有投资者对未来的看法是一致的。我们以后还可以从均衡定价的角度来进一步论证资本资产定价模型。这将在以后来介绍有关论述。

由通常的资本资产定价模型 (22) 可得

$$r_i - r_f = \beta_i(r_m - r_f) + \varepsilon_i, \tag{23}$$

其中 ε_i 是随机“干扰”，满足 $E[\varepsilon_i]=0$. 不但如此，还可由

$$\beta_i = \mathrm{Cov}[r_i, r_m]/\mathrm{Var}[r_m]$$

导得 $\mathrm{Cov}[r_m, \varepsilon_i]=0$. 因此，

$$\mathrm{Var}[r_i] = \beta^2\mathrm{Var}[r_m] + \mathrm{Var}[\varepsilon]. \tag{24}$$

这个等式经常被解释为：**证券收益的风险可分为两部分，一部分是市场带来的**系统风险，**另一部分是与市场无关的**非系统风险。对照这个数学表达式，我们应该对这两个名词有确切的理解。

这一资本资产定价模型 (CAPM) 的推导与前一讲中的推导很不一样。关于它与原来的推导的关系我们将在下面讨论。至于它的经济意义，正如我们以前已经说过，资本资产定价模型是刻划证券的超额收益与市场的超额收益之间的线性关系的。于是就存在这一模型能否在实际中被检验的问题。1976 年，Fama 在他的一本影响很大的著作 [4] 中指出，检验资本资产定价模型等价于检验市场组合的均值－方差有效性，而这样的检验是不可行的。Roll 在 1977 年发表的 [12] 中对资本资产定价模型的检验更是持批评态度。他们认为市场组合是否有效是无法检验的，因为它需要知道所有个体的资产状况，其中甚至包括未进入市场交易的资产，而检验近似的市场组合的有效性并不能说明市场组合是否有效。正因为有这样的论点，促使 Ross 的 APT 成为 CAPM 的替代物出现。APT 的一个好处在于它并不涉及市场组合的有效性问题。然而，Roll 在同一篇论文中实际上已指出我们这里的定理 4.3 的推论。这就是说，CAPM 的成立并不需要市场组合是有效的，而是只要能找到一个有效前沿上的组合，它 (与无风险证券一起) 就能起“市场组合”的作用，使得相应的 CAPM 成立。这就避免了市场组合是否有效的检验问题。

尽管如此，CAPM 还在继续为人们所应用。人们认为，“粗略地”应用 CAPM 还是有效的，其中市场组合通常就取各种市场指数 (Dow Jones 指数，S&P 指数等等)。由此得到的 β 可能没有严格的 CAPM 的意义，但是总能反映证券定价的信息及其风险状况。有一种简单的投资策略就是根据 CAPM 对证券的估价与证券的实际市场价的差别来决定证券的买卖。此外，在理论上，CAPM 更是成为某种基准，可用来讨论各种问题。

4.6　这一节中，我们将进一步讨论 Markowitz 证券组合选择理论与线性定价法则之间的关系。在前面的讨论中，我们总假定市场中有 n 种基本证券，甚至为求出证券组合的最优选择，还限定这 n 种证券的收益率协方差矩阵是正定的。但是这样的限制是不必要的。事实上，我们仍然可以在第二讲的基本框架中来讨论 Markowitz 理论。

回到第二讲中的基本框架，我们的讨论仍然可以从“未定权益 Hilbert 空间 $\mathcal{M}$”出发。在这个空间中有一个连续正齐次定价函数 $p\colon\mathcal{M}\to\mathbb{R}$. “收益集 R_1”仍然定义为

$$R_1=\{r\in\mathcal{M}\,|\,p(r)=1.\}$$

但需要假定它是仿射集，即对于任何 $r_1,r_2\in R_1$ 和任何实数 $\alpha\in\mathbb{R}$, 总有 $(1-\alpha)r_1+\alpha r_2\in R_1$. 然而，我们不需要假定 R_1 是有限维的，即它可能是由无限个“基本证券”所生成的收益集。不过由 p 是连续函数，这个仿射集一定是闭集。对于这样的收益集，我们仍然可以提出它的组合前沿的概念。注意到前面提出的二基金分离定理 4.2, “Markowitz 组合前沿存在条件”可以通过两个“前沿收益率” r_p 和 r_q 来提出：每个收益率 $r\in R_1$ 都不会改变由 r_p 和 r_q 生成的组合前沿。这首先要求：对于任何 $r\in R_1$, $r\neq r_p$, $r\neq r_q$, 存在 $\alpha_r\in\mathbb{R}$, 满足

$$\begin{aligned}E[r]&=(1-\alpha_r)E[r_p]+\alpha_rE[r_q]\\ \mathrm{Var}[r]&\geqslant\mathrm{Var}[(1-\alpha_r)r_p+\alpha_rr_q]\end{aligned}$$

不过这还并不是 r_p 和 r_q 生成对于 R_1 的元素的全体来说的组合前沿的充分条件。它还需要 r_p 和 r_q 都在它们生成的组合前沿上的条件：

$$\begin{aligned}E[r]&=(1-\alpha_r)E[r_p]+\alpha_rE[r_q]\\ \mathrm{Var}\left[\frac{r}{1-\alpha_r}-\frac{\alpha_rr_q}{1-\alpha_r}\right]&\geqslant\mathrm{Var}[r_p]\\ \mathrm{Var}\left[\frac{r}{\alpha_r}-\frac{(1-\alpha_r)r_p}{\alpha_r}\right]&\geqslant\mathrm{Var}[r_q]\end{aligned}$$

最后两个不等式与前面的一个不等式之间其实并不是完全独立的。我们实际上只需要其中的两个不等式。但在下面的定理 4.4 的条件 4) 中，我们还是都把它们列上。

定理 4.4 设未定权益空间 $\mathcal{M}$ 为方差有限的随机变量所构成的 Hilbert 空间. $p\colon\mathcal{M}\to\mathbb{R}$ 为 $\mathcal{M}$ 上的连续正齐次定价函数，即对于任何 $x\in\mathcal{M}$ 和任何 $\lambda>0$, 有 $p(\lambda x)=\lambda p(x)$. 同时假定 $p(\pm1_{\mathcal{M}})\neq0$, 这里 $1_{\mathcal{M}}\in\mathcal{M}$ 是无风险证券 1 (如果 $1\in\mathcal{M}$) 或无风险证券的模仿组合 (如果 $1\notin\mathcal{M}$)。定义

$$R_1=\{r\in\mathcal{M}\,|\,p(r)=1.\}$$

并假定它是 $\mathcal{M}$ 中的仿射集。那么下列四个命题等价:

1) 存在唯一的非零 $m\in\mathcal{M}$, 且 $E[r_m]=E[m]/E[m^2]\neq E[r_{1_{\mathcal{M}}}]=E[1_{\mathcal{M}}]/E[m]$, 使得对于任何 $x\in\mathcal{M}$, 有 $p(x)=E[mx]$;

2) (零 β-资本资产定价模型，zero-β-CAPM) 存在 $r_u \in R_1$, $E[r_u] \neq 0$, 使得对于任何 $r \in R_1$, 有

$$E[r] - E[r_v] = \frac{\mathrm{Cov}[r, r_u]}{\mathrm{Var}[r_u]}(E[r_u] - E[r_v]),$$

其中 $r_v \in R_1$, 满足 $E[r_v] \neq 0$, $\mathrm{Cov}[r_u, r_v] = 0$, $E[r_v] \neq E[r_u]$.

3) 存在 $r_p, r_q \in R_1$, $E[r_p] \neq E[r_q]$, 使得对于任何 $r \in R_1$, 存在实数 $\alpha \in \mathbb{R}$, 满足

$$\begin{aligned} E[r] &= (1-\alpha_r)E[r_p] + \alpha_r \mu_q, \\ \mathrm{Cov}[r, r_p] &= (1-\alpha_r)\mathrm{Var}[r_p] + \alpha_r \mathrm{Cov}[r_p, r_q], \\ \mathrm{Cov}[r, r_q] &= (1-\alpha_r)\mathrm{Cov}[r_p, r_q] + \alpha_r \mathrm{Var}[r_q]. \end{aligned}$$

4) 存在 $r_p, r_q \in R_1$, $E[r_p] \neq E[r_q]$, 使得对于任何 $r \in R_1$, 存在 $\alpha_r \in \mathbb{R}$, 满足

$$\begin{array}{rcl} E[r] & = & (1-\alpha_r)E[r_p] + \alpha_r E[r_q] \\ \mathrm{Var}[r] & \geqslant & \mathrm{Var}[(1-\alpha_r)r_p + \alpha_r r_q] \\ \mathrm{Var}\left[\dfrac{r}{1-\alpha_r} - \dfrac{\alpha_r r_q}{1-\alpha_r}\right] & \geqslant & \mathrm{Var}[r_p] \\ \mathrm{Var}\left[\dfrac{r}{\alpha_r} - \dfrac{(1-\alpha_r)r_p}{\alpha_r}\right] & \geqslant & \mathrm{Var}[r_q] \end{array}$$

证明 1) $\Rightarrow$ 4) (Hansen–Richard 正交分解定理) 设 1) 成立。由 1) 可知，r_m 与 $r_{1_{\mathcal{M}}}$ 不共线，即这两个向量张成 $\mathcal{M}$ 中的一个二维子空间 $\mathcal{M}_2$. 由 Hilbert 空间的正交分解定理，$\mathcal{M}$ 可分解为这个二维子空间与它的正交余空间的直和，即

$$\mathcal{M} = \mathcal{M}_2 + \mathcal{M}_2^{\perp}.$$

这就是说，对于任何 $x \in \mathcal{M}$, 存在实数 α_m^x 和 α_1^x 以及 $\varepsilon_x \in \mathcal{M}_2^{\perp}$, 使得

$$x = \alpha_m^x r_m + \alpha_1^x r_{1_{\mathcal{M}}} + \varepsilon_x, \quad E[r_m \varepsilon_x] = E[\varepsilon_x] = 0.$$

特别是，对于任何 $r \in R_1$, 存在实数 $\alpha_m, \alpha_r \in \mathbb{R}$, 使得

$$r = \alpha_m r_m + \alpha_r r_{1_{\mathcal{M}}} + \varepsilon_r, \quad E[r_m \varepsilon_r] = E[\varepsilon_r] = 0. \tag{25}$$

这是因为 $r, r_m, r_{1_{\mathcal{M}}} \in R_1$, 即 $E[mr] = E[mr_m] = E[mr_{1_{\mathcal{M}}}] = 1$, 同时，$E[m\varepsilon_r] = E[m]E[r_m\varepsilon_r] = 0$, 所以对上式的第一个等式两边乘上 m 再取数学期望，就可得到 $\alpha_m + \alpha_r = 1$, 即 $\alpha_m = 1 - \alpha_r$. 这样，由 (25) 可得，

$$\begin{aligned}
E[r] &= (1-\alpha_r)E[r_m] + \alpha_r E[r_{1_{\mathcal{M}}}] + E[\varepsilon_r] \\
&= (1-\alpha_r)E[r_m] + \alpha_r E[r_{1_{\mathcal{M}}}], \\
\mathrm{Var}[r] &= \mathrm{Var}[(1-\alpha_r)r_m + \alpha_r r_{1_{\mathcal{M}}}] + 2\mathrm{Cov}[(1-\alpha_r)r_m + \alpha_r r_{1_{\mathcal{M}}}, \varepsilon_r] + \mathrm{Var}[\varepsilon_r] \\
&= \mathrm{Var}[(1-\alpha_r)r_m + \alpha_r r_{1_{\mathcal{M}}}] + \mathrm{Var}[\varepsilon_r]
\end{aligned}$$

即 4) 中的等式和第一个不等式对于 $r_p = r_m, r_q = r_{1_{\mathcal{M}}}$ 成立。4) 中的后两个不等式的证明完全类似。

4) $\Rightarrow$ 3)　设 4) 成立。这其实是说明任何 $r \in R_1$ 都不能改变由 r_p 和 r_q 所生成的组合前沿。因此，3) 的成立可与定理 4.3 一样证明。但是我们这里给出另一个直接证明。事实上，设 $r \in R_1$ 满足

$$E[r] = (1-\alpha_r)E[r_p] + \alpha_r E[r_q].$$

令

$$r_\alpha = (1-\alpha_r)r_p + \alpha_r r_q.$$

那么对于任何实数 $\delta \in \mathcal{R}$, 我们都有

$$E[r_\alpha + \delta(r - r_\alpha)] = (1-\delta)E[r_\alpha] + \alpha E[r] = E[r] = E[r_\alpha].$$

因此，由 4) 成立，以及 R_1 是仿射集，$r_\alpha + \delta(r - r_\alpha) = (1-\delta)r_\alpha + \delta r \in R_1$, 可得

$$\mathrm{Var}[r_\alpha + \delta(r - r_\alpha)] \geqslant \mathrm{Var}[r_\alpha].$$

把左端展开，又得

$$\mathrm{Var}[r_\alpha] + 2\delta\mathrm{Cov}[r_\alpha, r - r_\alpha] + \delta^2\mathrm{Var}[r - r_\alpha] \geqslant \mathrm{Var}[r_\alpha].$$

如果取 $\delta > 0$, 那么由上式，并令 $\delta \to 0$, 可得

$$\mathrm{Cov}[r_\alpha, r - r_\alpha] \geqslant -\frac{\delta}{2}\mathrm{Var}[r - r_\alpha] \to 0.$$

但是取 $\delta < 0$, 由同样的推理，又可得到

$$\mathrm{Cov}[r_\alpha, r - r_\alpha] \leqslant -\frac{\delta}{2}\mathrm{Var}[r - r_\alpha] \to 0.$$

最后的结论是 $\mathrm{Cov}[r_\alpha, r - r_\alpha] = 0$, 即

$$\mathrm{Cov}[r_\alpha, r] = (1-\alpha_r)\mathrm{Cov}[r_p, r] + \alpha_r\mathrm{Cov}[r_q, r] = \mathrm{Var}[r_\alpha]. \tag{26}$$

用同样的方法同样可证明

$$\mathrm{Cov}\left[r_p, \frac{r-\alpha_r r_q}{1-\alpha_r} - r_p\right] = 0,$$

$$\mathrm{Cov}\left[r_q, \frac{r-(1-\alpha_r) r_p}{\alpha_r} - r_q\right] = 0.$$

而这两个等式即 3) 中的后两个等式。至于等式 (26)，只是这两个等式之和。

3) ⇒ 2)　设 3) 成立。如果 $\mathrm{Var}[r_p]\mathrm{Var}[r_q] = \mathrm{Cov}^2[r_p, r_q]$, 那么 r_p 和 r_q 完全相关。正如 4.2 节中所指出的那样，无风险利率将是 r_p 与 r_q 的组合。即存在实数 γ, 使得 $r_f = (1-\gamma)r_p + \gamma r_q$. 由于 $E[r_p] \neq E[r_q]$, 两者至少有一个不等于 r_f. 这时，不妨设 $E[r_p] \neq r_f$, 则可取 $r_u = r_p$, $r_v = r_f = (1-\gamma)r_p + \gamma r_q$.

如果 $\mathrm{Var}[r_p]\mathrm{Var}[r_q] > \mathrm{Cov}^2[r_p, r_q]$. 那么 $\mathrm{Var}[r_p]$ 和 $\mathrm{Var}[r_q]$ 中至少有一个不等于 $\mathrm{Cov}[r_p, r_q]$. 不妨设 $\mathrm{Var}[r_p] - \mathrm{Cov}[r_p, r_q] \neq 0$. 这时令 $r_u = r_p$, $r_v = (1-\gamma)r_p + \gamma r_q$, 其中 $\gamma \in \mathbb{R}$ 满足 $\mathrm{Cov}[r_u, r_v] = \mathrm{Cov}[r_p, (1-\gamma)r_p + \gamma r_q] = 0$, 即

$$\gamma = \frac{\mathrm{Var}[r_p]}{\mathrm{Var}[r_p] - \mathrm{Cov}[r_p, r_q]}.$$

由于 $\gamma \neq 0$, $E[r_p] \neq E[r_q]$, 同样有 $E[r_u] \neq E[r_v]$.

对这两种情形，我们都可把 $r_p = r_u$, $r_q = \frac{r_v}{\gamma} - \frac{(1-\gamma)r_u}{\gamma}$ 代入 3) 中的前两个等式，而得到

$$\begin{aligned}
E[r] &= (1-\alpha_r)r_u + \alpha_r\left(\frac{r_v}{\gamma} - \frac{(1-\gamma)r_u}{\gamma}\right) \\
&= \left(1-\alpha_r - \frac{\alpha_r(1-\gamma)}{\gamma}\right) r_u + \frac{\alpha_r r_v}{\gamma} \\
\mathrm{Cov}[r, r_u] &= (1-\alpha_r)\mathrm{Var}[r_u] + \alpha_r \mathrm{Cov}\left[r_u, \frac{r_v}{\gamma} - \frac{(1-\gamma)r_u}{\gamma}\right] \\
&= \left(1-\alpha_r - \frac{\alpha_r(1-\gamma)}{\gamma}\right)\mathrm{Var}[r_u]
\end{aligned}$$

由此立即可导得 2) 中要求的等式。

2) ⇒ 1)　见定理 2.3 的证明。 □

这样，我们就完全证明了 Markowitz 证券组合选择理论、资本资产定价模型和随机折现因子理论之间的等价性。这一结果是发人深思的。在从收益率出发的经典的 Markowtiz 理论中，尽管从表面上看来，线性

定价法则似乎应该已经隐身在背后。其实不然。事实上，该理论并没有禁止有不同收益率的多种无风险证券之类的套利现象存在。然而，如果要求有 Markowitz 意义下的组合前沿的存在，那么它就一定蕴含线性定价法则，以致不再可能出现类似的套利现象。换句话说，Markowitz 的组合前沿决定了一种唯一的线性定价法则。反过来也一样。

这里值得注意的是 Markowitz 的组合前沿只需要用两个前沿组合来确定。而由这两个前沿组合就能完全确定线性定价法则所需要的随机折现因子。作为与这两种理论等价的资本资产定价模型也同样如此。由此出发，在加上我们前面已经注意到的事实，可以断定，这些理论的数学其实都是“平面几何”，即它们都只涉及一个二维空间。这个空间是由随机折现因子 m 和无风险证券的模仿组合 $1_{\mathcal{M}}$ 所张成的。资本资产定价模型所涉及的 r_u 和 r_v 以及 Markowitz 证券组合前沿所涉及的 r_p 和 r_q, 都是这个二维空间中的两个不共线的向量。因此，知道其中的一对，就能求得其他两对。而未定权益 Hilbert 空间中的其他向量都被射影到这个二维空间上，以得到相应的“期望定价公式”。而对证券市场的定价问题在数学上最后都被归结为如何确定这个二维空间中的两个“基向量”。这就是 Black–Scholes 期权定价理论出现以前的金融经济学的核心所在。

4.7 在上面的讨论中，我们始终假定证券交易是允许卖空的。如果不允许卖空，即不允许 w_i 取负值，那么问题的求解不可能有 (10) 或 (15) 那样的明显的表达式。尤其是问题的求解也不再归结为线性方程组的求解。其实 Markowitz 的经典论文 [8] 倒是始终讨论不允许卖空的情形，并且为此提出**临界线** (critical line)**方法**。这时得到的有效前沿在不带无风险证券的情形将是若干直线段与双曲线段的拼接。由于我国股市至今还不允许卖空，因此讨论不允许卖空的证券组合选择问题更有现实意义。我们在这最后一小节对此给出一简要的讨论。

n 种证券的不允许卖空的 Markowitz 均值－方差组合选择问题为

$$\begin{array}{ll} \min & \sigma_w^2 = w^T V w = \sum_{i,j=1}^n V_{ij} w_i w_j \\ \text{s.t.} & w^T e = w_1 + w_2 + \cdots + w_n = 1 \\ & \mu_w = w^T \mu = w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n = \bar{\mu} \\ & w_i \geqslant 0, \quad i = 1, 2, \ldots, n. \end{array} \tag{27}$$

对于所有可能的组合收益 $\bar{\mu}$ 得到的问题 (27) 的解的全体就是这 n 种证券的 (不卖空) **极小风险组合**全体。n 种证券的有效组合一定是极小风险组合，但反之不然。与允许卖空的情形有所不同的是目前的极小风险组合的收益只可能在区间 $[\min_i \mu_i, \max_i \mu_i]$ 中来取，而有效前沿组合的收益更是只能在区间 $[\mu^n_{\min \sigma_w}, \max_i \mu_i]$ 中来取，这里 $\mu^n_{\min \sigma_w}$ 是对

应组合的最小风险 $\min\sigma_w$ 的最大收益，它是问题 (27) 中去掉第二个等式约束以后所得到的解组合所对应的最大收益。

问题 (27) 是一个带线性等式和不等式约束的二次凸规划问题。它也可用 Lagrange 乘子法 (或者说 Kuhn-Tucker 条件) 来求解，即 $\bar{w}=(\bar{w}_1,\bar{w}_2,\ldots,\bar{w}_n)^T$ 和 $\lambda_1,\lambda_2,\nu=(\nu_1,\ldots,\nu_n)^T$ 是其解和 Lagrange 乘子的充要条件为它们满足

$$\left\{\begin{array}{rcl} 2V\bar{w}-\lambda_1\mu-\lambda_2 e-\nu &=& 0, \\ \bar{w}_1+\bar{w}_2+\cdots+\bar{w}_n &=& 1, \\ \bar{w}_1\mu_1+\bar{w}_2\mu_2+\cdots+\bar{w}_n\mu_n &=& \bar{\mu}, \\ \nu_i\geqslant 0,\bar{w}_i\geqslant 0,\nu_i\bar{w}_i &=& 0, \\ i=1,2,\ldots,n. && \end{array}\right. \tag{28}$$

除了不允许卖空与允许卖空的两种情形以外，还有一部分证券允许卖空，另一部分证券不允许卖空的中间情形。不妨设 n 种证券的前 k' 种证券是允许卖空的。那么相应问题的解和 Lagrange 乘子仍然可以表达为上述形式，但是其中的 ν_i 要改成满足下列条件：

$$\nu_i=0,\quad i=1,2,\ldots,k';\qquad \nu_i\geqslant 0,\ \bar{w}_i\geqslant 0,\ \nu_i\bar{w}_i=0,\quad i=k'+1,\ldots,n.$$

为了表达方便，我们把 n 中证券所组成的集合表示为 S_n. $S_{k'}$ 表示其前 k' 种证券的集合。 S_J 则表示 S_n 的一个子集。

由于 (28) 的每一个解 $\bar{w}$ 总有一些分量 $\bar{w}_i>0$, 从而其相应的 ν_i 为零，于是我们立即可以得到下列命题：

命题 4.1 设 $\bar{w}=(\bar{w}_1,\bar{w}_2,\ldots,\bar{w}_n)^T$ 为问题 (27) 对于收益为 $\bar{\mu}$ 的一个解，且在其充要条件 (28) 的表达中，满足

$$\bar{w}_i\left\{\begin{array}{ll} >0, & i\in S_J=\{j_1,j_2,\ldots,j_l\}\subset S_n, \\ =0 & i\notin S_J. \end{array}\right. \tag{29}$$

那么 $\bar{w}$ 也是对 S_J 中的证券允许卖空，而对 $S_n\setminus S_J$ 不允许卖空的极小风险组合。特别是，$\bar{w}'=(\bar{w}_{j_1},\bar{w}_{j_2},\ldots,\bar{w}_{j_l})^T$ 是 S_J 上的允许卖空的 Markowitz 问题 (5) 的解。

我们不妨称满足条件 (29) 的 S_J 为组合 $\bar{w}$ 的**指标集**。由这一定义和 (6)–(8) 的线性性，我们立即可得下列命题。

命题 4.2 设 $\bar{w}^m=(\bar{w}_1^m,\bar{w}_2^m,\ldots,\bar{w}_n^m)^T$ 为问题 (5) 对于收益为 $\bar{\mu}^m$ 的解，$m=1,2$, 且它们都以 S_J 为指标集。那么对于任何 $\alpha\in(0,1)$, $\bar{w}^\alpha=(1-\alpha)\bar{w}^1+\alpha\bar{w}^2$ 是问题 (27) 对于收益为 $\bar{\mu}^\alpha=(1-\alpha)\bar{\mu}^1+\alpha\bar{\mu}^2$ 的解，并且 $\bar{w}^\alpha$ 也以 S_J 为指标集。

这样一来，在 S_n 中至少有两种收益不同的证券时，问题 (27) 可能有解的收益 $\bar{\mu}$ 可分为两类, 一类形成 $[\min_i \mu_i, \max_i \mu_i]$ 中的一个区间，在这个区间上，对于每一个 $\bar{\mu}$，都存在有同样指标集 S_J 的问题 (27) 的解，以致它们都是同一个指标集 S_J 上的允许卖空的问题 (5) 的解。这个 S_J 至少包含两种收益不同的证券，否则 S_J 上的允许卖空问题的解不可能对 $\bar{\mu}$ 在一个区间上的值都存在。另一类则形不成区间，但对应的解也是某个指标集的允许卖空问题的解。因为 S_n 只有有限个子集 S_J, 所以 $[\min_i \mu_i, \max_i \mu_i]$ 一定是有限个这样的区间和有限个点的并集。即存在 $\bar{\mu}^l \in [\min_i \mu_i, \max_i \mu_i]$, $l = 1, \ldots, p$, 满足

$$\min_i \mu_i = \bar{\mu}^1 < \bar{\mu}^2 < \cdots < \bar{\mu}^p = \max_i \mu_i,$$

使得在区间 $(\bar{\mu}^l, \bar{\mu}^{l+1})$ 中，对应这一区间中的 $\bar{\mu}$ 的解可以有同样的指标集 S_J. 也就是说，对于这一区间中的 $\bar{\mu}$ 来说，所对应的解也一定是 S_J 上的允许卖空的问题的 (27) 的解。至于对于区间的端点 $\bar{\mu}^l$ 和 $\bar{\mu}^{l+1}$ 来说，它们对应的前沿组合被 Markowitz 称为**拐角组合** (corner portfolio)，具有很特殊的意义。

我们从前面的讨论中已经知道，在允许卖空的情形下，如果 S_n 的组合前沿非空，那么它或是一个孤立点 (所有 μ_i 都相等)，或是双曲线的右半部 (S_n 的组合中不包括无风险资产)，或是起点在纵轴上的两条斜率相反的射线 (S_n 的组合中包含无风险资产)。这样一来，在上述以 S_J 为指标集的解所对应的 $\bar{\mu}$ 的区间中，其不允许卖空的组合前沿必定与 S_J 上允许卖空的组合前沿重合，即它或者是一段双曲线，或者是一段直线。而以 $\bar{\mu}$ 为自变量，问题 (27) 的值 σ 为因变量的函数 $\sigma(\bar{\mu})$ 的整个图象将是这种双曲线段或直线段拼接而成的 $[\min_i \mu_i, \max_i \mu_i]$ 上的连续凸函数的图象。这里连续性和凸性不难直接证明。尽管利用指标集来对区间 $[\min_i \mu_i, \max_i \mu_i]$ 进行上述分割的方法可能不唯一，但形成的函数 $\sigma(\bar{\mu})$ 是没有区别的。综上所述，我们就得到下列定理：

定理 4.5 设 S_n 为一个非空证券集。那么它的不允许卖空条件下的组合前沿非空，并且它或是一个孤立点 (所有 μ_i 都相等)，或是有限段双曲线段或直线段的联结。问题 (27) 的值

$$\sigma(\bar{\mu}) = \min_{w \in \mathbb{R}^n_+, w^T e = 1, w^T \mu = \bar{\mu}} \left(w^T V w\right)^{1/2},$$

关于收益 $\bar{\mu}$ 的函数 σ 在 $(\bar{\mu}, \sigma_1)$ 平面上的图象是由有限段双曲线段或直线段联结而成的连续凸函数图象；从而当 $\bar{\mu} \geqslant \mu_{\min \sigma_w}$ 时，其反函数图象 (即有效前沿) 在 $(\sigma_w, \bar{\mu})$-平面上形成一个有限闭区间上的连续凹函数的图象。

定理 4.5 是对于组合前沿而言的，对于有效前沿来说，结论类似。但有效前沿归结为一点时并不能导得所有 μ_i 相等，还有可能是最小风险组合的收益最大。这样，我们对不允许卖空的均值－方差组合选择问题的解也就有一个大致的了解。需要引起注意的是，由于 (28) 不再是线性方程组，不允许卖空的二基金分离定理就不再成立。因此，上述思路的资本资产定价模型也不再成立。上一节中所提出的三种理论的等价性也不再成立。这尤其是说明，如果我们希望通过不允许卖空情形的组合前沿来对证券定价，那么这种定价法则将不再是线性定价法则。这一问题值得人们进一步探讨。上述的一些事实似乎并未为一些研究者所注意到。事实上，我们经常可以看到有关中国股市的 CAPM 的研究。不过，我们在后面的两讲中又可以看到，在一般经济均衡框架中的 CAPM, 会得出投资者一定会选择不卖空的投资策略。于是在一般均衡框架下， CAPM 在允许卖空与不允许卖空两种情形之间的区别又将变得不明显。

说明　Markowitz 证券组合选择理论在今天作为现代金融经济学的起点，已经变为经典理论。每一本这方面的教科书都要相当详尽地讲述它。例如，Huang–Linzenberger [6] 和 Ingersoll [8] 都有相当大的篇幅介绍 Markowitz 理论。Jarrow [9] 的观点稍有不同，它先用一般均衡框架讲 CAPM, 再提出均值－方差有效的概念来介绍 Markowitz 理论。此外，每本《投资学》教科书也一定要介绍 Markowitz 理论。 Markowitz 理论在数学上并不很困难，不过其表达式有点繁琐。这里我们采用 Constantinides–Malliaris [3] 的矩阵表达方式，使形式显得比较简洁。但是需要对矩阵运算比较熟悉。本章中除用到线性代数、概率论和微分学的一些知识都可以在有关教科书中找到外，还用到一些数学规划的知识。这些知识也可在任何一本数学规划 (最优化方法) 方面的教科书中找到。一些简要的并且不同于通常的叙述见我们的“附录 2”。

上面提到的所有教科书在介绍 Markowitz 理论时都是假定允许卖空的。其实无论在 Markowitz 的开创性论文 [10] 中，还是在他后来发表的专著 [11] 中，都是假定不允许卖空的。在允许卖空的情形下，可以利用线性方程组求解来展开理论，问题比较简单。同时，对后来的金融理论影响也更大。而在不允许卖空的情形下，数学上有点困难。我们在最后一节所介绍的一些不允许卖空情形的结果，取自我们的工作 [17]。这里所提出的是一种简洁的论证有效前沿的方法。

正如我们已经提到过，资本资产定价模型是 Sharpe [14], Lintner [9] 和 Mossin [12] 分别独立提出的。他们实际上都是把它与二基金分离定理相联系而导出的。二基金分离定理通常认为出于 Tobin [15] 和 Lintner [9]. 目前许多教科书中采用的资本资产定价模型的证明来自 Sharpe [14] 的证明。这一证明，正如我们在附录中所介绍的，严格地说，在数学上是不能令人满意的。在 Constantinides–Malliaris [3] 和 Huang–Linzenberger [6] 中则利用所谓“正交前沿组合”的概念来直接导出一般的 Black [1] 的“零 β 资本资产定价模型”。我们的证明来自我们的工作 [16], 在那里我们还证明了更一般的有关“增加若干证券不改变原来的有效前沿”的充要条件。在 [17] 中，我们还对不允许卖空的情形讨论了类似的问题。至于 Markowitz 有效前沿

存在，CAPM 和线性定价函数的存在之间的关系在 Cochrane [2] 中有一些类似的讨论，但与我们这里提出的有很大不同。定理 4.4 中的 1) $\Rightarrow$ 4) 称为 Hansen–Richard 分解定理，它是 1987 年发表的 Hansen–Richard [5] 中的结果。

关于资本资产定价模型的检验问题，在 Huang–Linzenberger [6] 中有专门一章进行讨论。

思考题和习题

1. 设 $y^p = \theta_1^p x_1 + \theta_2^p x_2 + \cdots + \theta_n^p x_n$, $p = 1, 2$ 为两个证券组合的未来价格。试指出证券组合 $y = y^1 + y^2$ 的收益率 r_y 与 y^1、y^2 的收益率 r_{y^1} 和 r_{y^2} 之间的关系。特别是指出，它们作为 n 种证券的收益率的组合的相互关系。

2. 设 p 和 q 为两种证券，且它们的收益率 r_p 和 r_q 满足 $E[r_p] = E[r_q]$. 那么这两种证券的组合前沿是什么？证明，证券 p 是这两种证券的前沿组合的充要条件为 $\mathrm{Cov}[r_p, r_q] = \mathrm{Var}[r_p]$.

3. 如果 n 种证券的期望收益率都相等，但收益率方差可能不同。同时，它们又都是互不相关的。试求出其有效组合，并写出其最小方差的表达式。

4. 设证券 p 和证券 q 有同样的收益 (期望收益率) 和风险 (收益率标准差)。试讨论这两种证券的组合，并指出其前沿组合与协方差 $\mathrm{Cov}[r_p, r_q]$ 之间的关系。

5. 对于任意的证券集，有效前沿是否一定存在？试讨论两种有同样的风险、不同的收益的证券的组合。当它们的相关系数为 1 时，它们的有效前沿存在吗？

6. 如果在证券集中有两种收益不同的无风险证券，那么它们对该证券集的有效前沿有什么影响？

7. 设证券组合 $w_p = (w_p^1, w_p^2, \ldots, w_p^n)^T$, $w_q = (w_q^1, w_p^2, \ldots, w_q^n)^T$. 那么它们的收益率的协方差 $\mathrm{Cov}[r_p, r_q]$ 该怎样用 n 种证券的协方差矩阵来表示？试由此证明，如果 w_p 是极小风险组合，并且不是总体极小风险组合，那么一定存在另一个极小风险组合 w_q, 使得 $\mathrm{Cov}[r_p, r_q] = 0$.

8. 什么是资本资产定价模型中的市场组合？怎样来论证市场组合是有效组合，并且它不包含无风险证券？

9. 什么是资本市场线？什么是证券市场线？请用证券的收益率来写出它们的表达式，并指出它们的不同含义及相互关系。

10. 什么是证券收益的系统风险和非系统风险？能不能把系统风险称为“市场风险”？

11. 怎样看待资本资产定价模型与 Markowitz 证券组合选择理论之间的关系？

12. 本讲导出的资本资产定价模型与上一讲中的资本资产定价模型有什么不同？它们之间有什么关系？

13. 怎样看待 Markowitz 证券组合选择理论、资本资产定价模型和线性定价法则之间的等价关系？怎样认识这里涉及的数学本质上是一种“平面几何”？

参考文献

[**1**] Black， F., 1972, Capital market equilibrium with restricted borrowing, *Journal of Business,* **45:** 444–454.

[**2**] Cochrane, J. H., 2001, Asset Pricing, Princeton University Press, Princeton, N. J..

[**3**] Constantinides, G. M. & A. G. Malliaris, 1995, Portfolio theory, in R. A. Jarrow, V. Maksimovic & W. T. Ziemba (eds), *Finance*, Elsevier, Amsterdam, 1–30.

[**4**] Fama, E. F., 1976, Foundations of Finance, Basic Books, New York, NY.

[**5**] Hansen, L. P. & S. F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica*, **55:** 587–614.

[**6**] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[**7**] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[**8**] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[**9**] Lintner, J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, **47:** 13–37.

[10] Markowitz, H., 1952, Portfolio selection, *Journal of Finance*, **7:** 77–91.

[11] Markowitz, H., 1959, 1991 Second ed., Portfolio Selection: Efficient Diversification of Investment, Basil Blackwell, Cambridge.

[12] Mossin, J., 1966, Equilibrium in a capital asset market, *Econometrica*, **34:** 768–783.

[13] Roll, R., 1977, A critique of the asset pricig theory's tests: Part 1, *J. Financ. Econ.*, **4:** 129–176.

[14] Sharpe, W., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, **19:** 425–442.

[15] Tobin, J., 1958, Liquidity preference as bihavior toward risk, *Rev. Econ. Studies*, **25:** 65–86.

[16] 史树中、杨杰，2002, 证券组合选择的有效子集，《应用数学学报》**55**: 176–186.

[17] 史树中、杨杰，2003, 不允许卖空证券组合选择的有效子集，《应用数学学报》**56**:.

附录 1：资本资产定价模型的 Sharpe 证明

Sharpe 的证明基于这样的思想：对于任何市场中的证券 (或证券组合) i, 它与市场组合 m 的组合所形成的风险－收益双曲线必定与资本市场线相切于市场组合所对应的点 (σ_m, μ_m) 上。由 i 和 m 所生成的风险－收益曲线可用参数形式来表达：

$$\begin{cases} \sigma = \sigma(\alpha) = ((1-\alpha)^2 \mathrm{Var}[r_m] + 2\alpha(1-\alpha)\mathrm{Cov}[r_i, r_m] + \alpha^2 \mathrm{Var}[r_i])^{1/2} \\ \mu = \mu(\alpha) = (1-\alpha)E[r_m] + \alpha E[r_i] \end{cases}$$

这里 r_m 和 r_i 分别是市场组合 m 与证券 (组合) i 的收益率。资本市场线则可表达为 $\mu - r_f = \gamma\sigma$, 其中 r_f 是无风险利率。由于市场组合对应的点 $(\sigma_m, \mu_m) = (\mathrm{Var}^{1/2}[r_m], E[r_m])$ 在资本市场线上，我们有

$$\gamma = \frac{E[r_m] - r_f}{\mathrm{Var}^{1/2}[r_m]}.$$

另一方面，我们又在上述风险－收益曲线上有 $\sigma_m = \sigma(0)$ 和 $\mu_m = \mu(0)$, 而由于资本市场线与上述风险－市场曲线相切于 (σ_m, μ_m)，故该曲线在相应点的斜率也等于 γ, 即

$$\gamma = \frac{\mu'(0)}{\sigma'(0)} = \frac{(E[r_m] - E[r_i])\mathrm{Var}^{1/2}[r_m]}{\mathrm{Var}[r_m] - \mathrm{Cov}[r_m, r_i]}.$$

由这两个 γ 的表达式，立即可得资本资产定价模型：

$$E[r_i]-r_f=\frac{\mathrm{Cov}[r_m,r_i]}{\mathrm{Var}[r_m]}(E[r_m]-r_f).$$

附录 2：数学预备知识 2

1. 凸集、凸函数和凹函数： n 维向量空间 $\mathbb{R}^n$ 中的集合 K 称为**凸集**是指

$$\forall x,y\in K,\forall\lambda\in[0,1],\qquad (1-\lambda)x+\lambda y\in K.$$

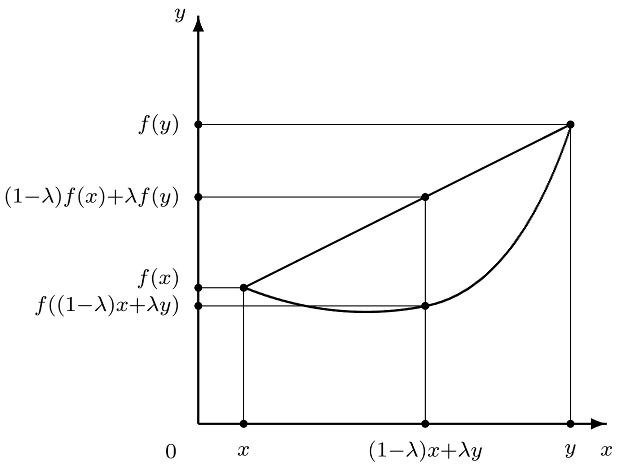
其几何意义是集合 K 中的任何两点的连接线段都在 K 中。这个概念也能推广到一般的 (无限维) 向量空间上。

设 $f\colon K\to\mathbb{R}$, 其中 $K\subset\mathbb{R}^n$ 是一个凸集。如果对于任何 $x,y\in K$, $\lambda\in(0,1)$, 有

$$f((1-\lambda)x+\lambda y)\leqslant(1-\lambda)f(x)+\lambda f(y),$$

那么 f 称为凸集 K 上的**凸函数**。$u\colon K\to\mathbb{R}$ 称为 K 上的**凹函数**，是指 $f=-u$ 是凸函数。如果涉及的不等式都是严格不等式，那么相应的概念都可加上“严格”。即严格凸函数或严格凹函数等等。

对于单变量的凸函数 $f\colon(a,b)\to\mathbb{R}$, 上述不等式的含义可从以下的图看出。



凸函数的图象

如果 $b > y > z > x > a$, 那么存在 $\lambda \in (0,1)$, 使得 $z = (1-\lambda)x + \lambda y$, $\lambda = (z-x)/(y-x)$. 于是

$$f(z) \leqslant \frac{y-z}{y-x} f(x) + \frac{z-x}{y-x} f(y).$$

由此可得

$$\frac{f(x)-f(z)}{x-z} \leqslant \frac{f(y)-f(z)}{y-z}.$$

如果 f 可导，那么分别令 $z \to x$, $z \to y$, 可得

$$\forall x < y, \qquad f'(x) \leqslant \frac{f(y)-f(x)}{y-x} \leqslant f'(y).$$

这个条件也是可导的 f 为凸函数的充分条件 (证明要用中值定理)。这样，可导的凸函数 f 可定义为其导数 f' 递增的函数。二阶可导的凸函数 f 可定义为其二阶导数 $f'' \geqslant 0$ 的函数。同样，可导的凹函数 u 可定义为其导数 u' 递减的函数。二阶可导的凹函数 u 可定义为其二阶导数 $u'' \leqslant 0$ 的函数。

这些结果都可推广到多变量 $x = (x_1, x_2, \ldots, x_n)^T$ 情形。这时一阶导数

$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)^T,$$

为 f 在 $x \in K$ 处的梯度向量。 f' 在 x 上单调是指对于任何 $x' \in K$, $(f'(x') - f'(x))^T (x' - x) \geqslant 0$. 与此同时，设 $x, y \in K$, $h = y - x$, 那么对于 K 上的凸函数还有

$$f(y) - f(x) = f(x+h) - f(x) \geqslant (f'(x))^T h = (f'(x))^T (y-x).$$

而二阶导数

$$f''(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x) & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dfrac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dfrac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

为 f 在 $x \in K$ 处的 Hesse 矩阵。f'' 在 x 上非负是指该矩阵非负定。凹函数的情形类推。

Markowtitz 问题中的目标函数的 Hesse 矩阵就是证券收益率的协方差矩阵 V, 它是非负定的，因此，它是凸函数。其约束集合是两个线性等式决定的凸集。

2. **欧氏空间中集合的正交性和线性方程组理论：**

我们在前面已经证明，对于 Hilbert 空间 (包括有限维的欧氏空间) X 的每个闭子空间 Y, 都可以有下列正交分解，

$$X = Y \oplus Y^{\perp},$$

即对于任何 $x \in X$, 存在唯一的 $y_x \in Y$ 和 $y_x^{\perp} \in Y^{\perp}$, 使得 $x = y_x + y_x^{\perp}$.

由此可导得：**如果有一个向量 z 有这样的性质：对于任何与 Y 正交的向量，也一定与 z 正交，那么 z 一定是 Y 的元素。**事实上，这是因为 z 与 $Y^{\perp}$ 中的所有元素都正交，所以它的分解式中不可能有非零的 $Y^{\perp}$ 成分。

n 个变量 $x_1, x_2, \ldots, x_n$ 的齐次线性方程

$$a_1 x_n + a_2 x_n + \cdots + a_n x_n = 0$$

可以看作求一个 n 维向量 $(x_1, x_2, \ldots, x_n)$ 使得它与 n 维向量 $(a_1, a_2, \ldots, a_n)$ 正交。因此，前面的结论可用于齐次线性方程组理论。由此可得到：**如果 k 个齐次线性方程形成的方程组的解一定也是第 $k+1$ 个齐次线性方程的解，那么这第 $k+1$ 个方程一定是前 k 个方程的线性组合。**

证明概要：设 k 个线性齐次方程为：

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ \cdots\cdots\cdots\cdots\cdots & \cdots & \cdots \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n & = & 0 \end{array}$$

它可看作求 n 维向量 $x = (x_1, \ldots, x_n)$, 使得它与 k 个 n 维向量 $a^i = (a_{i1}, \ldots, a_{ik})$, $i = 1, \ldots, k$, 都正交。令这 k 个向量张成的向量空间为 Y, 那么满足方程的 x 的全体就形成 $Y^{\perp}$. 如果把第 $k+1$ 个线性齐次方程也看作一个 n 维向量 $a^{k+1} = (a_{k+1,1}, \ldots, a_{k+1,n})$ 与 x 正交的表达式，那么命题的条件是说 a^{k+1} 与 $Y^{\perp}$ 中的所有元素都正交。因此，它一定是 Y 中的元素，即它可表示成前 k 个向量的线性组合。

上述结果中的方程的齐次性条件可以去掉，即**如果 k 个线性方程形成的方程组的解一定也是第 $k+1$ 个线性方程的解，那么这第 $k+1$ 个方程一定是前 k 个方程的线性组合。**事实上，设 k 个线性方程为：

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \cdots\cdots\cdots\cdots\cdots & \cdots & \cdots \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n & = & b_n \end{array}$$

不妨设它有解为 $x^0=(x_1^0,\ldots,x_n^0)$, 那么令 $y=x-x^0=(y_1,\ldots,y_n)$, 原方程变为 k 个线性齐次方程：

$$\begin{array}{rcl} a_{11}y_1+a_{12}y_2+\cdots+a_{1n}y_n & = & 0\\ a_{21}y_1+a_{22}y_2+\cdots+a_{2n}y_n & = & 0\\ \cdots\cdots\cdots\cdots\cdots & \cdots & \cdots\\ a_{k1}y_1+a_{k2}y_2+\cdots+a_{kn}y_n & = & 0 \end{array}$$

因此，非齐次问题可归结为齐次问题来讨论。

3. Markowitz 问题上的应用：

考虑 Markowitz 问题：

$$\left\{\begin{array}{ll} \min & V(w)=\sigma_w^2=w^TVw=\sum_{i,j=1}^n V_{ij}w_iw_j\\ \text{s.t.} & w^Te=w_1+w_2+\cdots+w_n=1\\ & \mu_w=w^T\mu=w_1\mu_1+w_2\mu_2+\cdots+w_n\mu_n=\bar{\mu} \end{array}\right. \tag{1}$$

令

$$H=\{w\in\mathbb{R}^n\,|\,w^Te=1,w^T\mu=\bar{\mu}.\}$$

为问题 (1) 的约束集。如果 $\bar{w}$ 是问题 (1) 的解，那么对于 $w\in H$, 都有 $V(w)\geqslant V(\bar{w})$. 由于 H 是一个凸集，故对于任何 $t\in(0,1)$, $(1-t)\bar{w}+tw\in H$, 从而我们有

$$V(\bar{w}+t(w-\bar{w}))-V(\bar{w})\geqslant 0. \tag{2}$$

令 $h=w-\bar{w}$. 则由 w 和 $\bar{w}$ 都是 H 的元素， h 满足齐次线性方程组

$$\begin{array}{rcl} h_1+h_2+\cdots+h_n & = & 0\\ \mu_1h_1+\mu_2h_2+\cdots+\mu_nh_n & = & 0 \end{array} \tag{3}$$

同时，每个满足上述条件的 h 也一定对应一个 $w=\bar{w}+h\in H$. 由 (2) 可得

$$\lim_{t\to0+}\frac{V(\bar{w}+th)-V(\bar{w})}{t}\geqslant 0,$$

上式左端是函数 V 沿方向 h 的方向导数，它等于

$$\frac{\partial V(\bar{w})}{\partial w_1}h_1+\frac{\partial V(\bar{w})}{\partial w_2}h_2+\cdots+\frac{\partial V(\bar{w})}{\partial w_n}h_n.$$

上式说明，它对于任何满足 (3) 的 h 非负。但是如果 h 满足 (3), 那么 $-h$ 也满足 (3). 因此，实际上对于任何满足 (3) 的 h 必定有

$$\frac{\partial V(\bar{w})}{\partial w_1}h_1+\frac{\partial V(\bar{w})}{\partial w_2}h_2+\cdots+\frac{\partial V(\bar{w})}{\partial w_n}h_n=0. \tag{4}$$

利用上节中的结果，我们可知，这个对于 h 的方程一定是前两个方程的线性组合，即存在实数 λ_1, λ_2, 使得

$$\begin{array}{rcl}
\dfrac{\partial V(\bar{w})}{\partial w_1} & = & \lambda_1\mu_1 + \lambda_2 \\
\dfrac{\partial V(\bar{w})}{\partial w_2} & = & \lambda_1\mu_2 + \lambda_2 \\
\cdots\cdots & \cdots & \cdots\cdots\cdots \\
\dfrac{\partial V(\bar{w})}{\partial w_n} & = & \lambda_1\mu_n + \lambda_2
\end{array}$$

即

$$\begin{array}{rcl}
V_{11}\bar{w}_1 + V_{12}\bar{w}_2 + \cdots + V_{1n}\bar{w}_n - \dfrac{1}{2}\lambda_1\mu_1 - \dfrac{1}{2}\lambda_2 & = & 0 \\
V_{21}\bar{w}_1 + V_{22}\bar{w}_2 + \cdots + V_{2n}\bar{w}_n - \dfrac{1}{2}\lambda_1\mu_2 - \dfrac{1}{2}\lambda_2 & = & 0 \\
\cdots\cdots\cdots\cdots & \cdots & \cdots \\
V_{n1}\bar{w}_1 + V_{n2}\bar{w}_2 + \cdots + V_{nn}\bar{w}_n - \dfrac{1}{2}\lambda_1\mu_n - \dfrac{1}{2}\lambda_2 & = & 0
\end{array}$$

或

$$2V\bar{w} - \lambda_1\mu - \lambda_2 e = 0. \tag{5}$$

反之，如果存在 λ_1, λ_2, 使得 $\bar{w} \in H$ 满足 (5)，那么它对于任何满足 (3) 的 h 必定有 (4) 成立。由于 V 是凸函数，故对于任何满足 (3) 的 h 必定有

$$V(\bar{w} + h) - V(\bar{w}) \geqslant \lim_{t\to 0+} \frac{V(\bar{w} + th) - V(\bar{w})}{t} = 0,$$

即对于任何满足 (1) 的约束条件的 w 有 $V(w) \geqslant V(\bar{w})$. 这就是说，问题 (1) 的约束条件和 (5) 两者在一起，也是问题 (1) 的解的充分条件。

4. 不允许卖空的 Markowitz 问题：

现在考虑不允许卖空的 Markowitz 问题：

$$\left\{\begin{array}{ll}
\min & V(w) = \sigma_w^2 = w^T V w = \sum_{i,j=1}^n V_{ij} w_i w_j \\
\text{s.t.} & w^T e = w_1 + w_2 + \cdots + w_n = 1 \\
 & \mu_w = w^T \mu = w_1\mu_1 + w_2\mu_2 + \cdots + w_n\mu_n = \bar{\mu} \\
 & w_1, w_2 \ldots, w_n \geqslant 0
\end{array}\right. \tag{6}$$

这里仅当 $\bar{\mu} \in [\min_i \mu, \max_i \mu]$ 时，问题才有解。

仍然令 $H = \{w \in \mathbb{R}^n \,|\, w^T e = 1, w^T\mu = \bar{\mu}.\}$. 我们不妨还假定：

存在 $\tilde{w} \in H$, *使得* $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n > 0$.

这个条件当 $\bar{\mu} \neq \min_i \mu$ 或 $\max_i \mu$ 时总是成立的。这是因为对于任何第 i 种证券，它总能与收益最大的证券或收益最小的证券一起构成一个不允许卖空的组合，使得组合收益恰好等于 $\bar{\mu}$. 然后，把这些组合再放在一起构成一个不允许卖空的每个成分的比例都为正的组合，该组合所对应的 $\tilde{w}$ 就会满足上述条件。至于 $\bar{\mu} = \min_i \mu$ 或 $\max_i \mu$ 的情形相对来说比较简单，可用下面的类似方法进行讨论。

不难看出，各分量非负的向量 $\bar{w} \in H$ 是问题 (27) 的解的充分必要条件为

$$\forall w \in H, \qquad \max\{V(w) - V(\bar{w}), -w_1, \ldots, -w_n\} \geqslant 0. \tag{7}$$

因此，集合

$$\begin{aligned} A = \{x = (x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} \quad | \quad & \exists w \in H, \text{使得} \\ & V(w) - V(\bar{w}) \leqslant x_0, -w_i \leqslant x_i, i = 1, \ldots, n.\} \end{aligned}$$

与集合

$$B = \{y = (y_0, y_1, \ldots, y_n) \in \mathbb{R}^{n+1} \,|\, y_j < 0, j = 0, 1, \ldots, n.\}$$

不相交。B 显然是凸集。而对于任何 $x^1, x^2 \in A$, 存在 $w^1, w^2 \in H$, 使得

$$V(w^p) - V(\bar{w}) \leqslant x_0^p, \quad -w_i^p \leqslant x_i^p, \quad p = 1, 2;\ i = 1, \ldots, n.$$

因此，对于任何 $\lambda \in (0, 1)$,

$$\begin{array}{rcl} V((1-\lambda)w^1 + \lambda w^2) - V(\bar{w}) & \leqslant & (1-\lambda)(V(w^1) - V(\bar{w})) + \lambda(V(w^2) - V(\bar{w})) \\ & \leqslant & (1-\lambda)x^1 + \lambda x^2; \\ -[(1-\lambda)w_i^1 + \lambda w_i^2] & \leqslant & (1-\lambda)x_i^1 + \lambda x_i^2, \quad i = 1, \ldots, n, \end{array}$$

这里用到凸函数 V 的性质，即它满足

$$V((1-\lambda)w^1 + \lambda w^2) \leqslant (1-\lambda)V(w^1) + \lambda V(w^2).$$

这说明 $(1-\lambda)x^1 + \lambda x^2 \in A$, 以致 A 也是凸集。这样，我们就可应用凸集分离定理：存在 $\nu_0, \nu_1, \ldots, \nu_n$ 不全为零，使得

$$\forall x \in A, \forall y \in B, \qquad \nu_0 x_0 + \nu_1 x_1 + \cdots + \nu_n x_n \geqslant \nu_0 y_0 + \nu_1 y_1 + \cdots + \nu_n y_n.$$

这一不等式的右端可任意接近于零，因而,

$$\forall x \in A, \qquad \nu_0 x_0 + \nu_1 x_1 + \cdots + \nu_n x_n \geqslant 0.$$

同时，由于 $x_0, x_1, \ldots, x_n$ 都可以任意大，故必须有 $\nu_0, \nu_1, \ldots, \nu_n \geqslant 0$. 另一方面，前面的假定还可导出 $\nu_0 > 0$. 否则，在上式中取 $x_i = w_i$, 可得到对于任何 $w \in H$, 有

$$\nu_1 w_1 + \nu_2 w_2 + \cdots + \nu_n w_n = 0.$$

根据假定，H 中有各个分量都为正的向量 $\tilde{w}$. 这使得上式仅当所有 $\nu_i = 0$ 时才有可能。与它们不全为零矛盾。这样，必须有 $\nu_0 > 0$. 不妨取 $\nu_0 = 1$. 根据集合 A 的定义，由此导得

$$\forall w \in H, \qquad V(w) - \nu_1 w_1 - \nu_2 w_2 - \cdots \nu_n w_n \geqslant V(\bar{w}).$$

把 $\bar{w}$ 代到左端，还能得到

$$-\nu_1 \bar{w}_1 - \nu_2 \bar{w}_2 - \cdots - \nu_n \bar{w}_n \geqslant 0.$$

由于所有的 $\nu_i \geqslant 0$, $\bar{w}_i \geqslant 0$, 这仅当

$$\nu_i \bar{w}_i = 0, \qquad i = 1, \ldots, n,$$

时才有可能。因此，最后有：$\bar{w} = (\bar{w}_1, \ldots, \bar{w}_n) \in H$ **是问题** (27) **的解的充分必要条件为：存在** $\nu_1, \ldots, \nu_n \geqslant 0$, **使得**

$$\forall w \in H, \qquad V(w) - \nu_1 w_1 - \cdots - \nu_n w_n \geqslant V(\bar{w}).$$

$$\bar{w}_i \geqslant 0, \qquad \nu_i \bar{w}_i = 0, \quad i = 1, \ldots, n.$$

前面证明的是条件的必要性。而条件的充分性几乎是显然的。

我们再把 $V(w, \nu) = V(w) - \nu_1 w_1 - \cdots - \nu_n w_n$ 来代替上一节中的 $V(w)$, 并进行完全一样的讨论，结果就得到：$\bar{w} = (\bar{w}_1, \ldots, \bar{w}_n) \in H$ **是问题** (27) **的解的充分必要条件为：存在** $\nu_1, \ldots, \nu_n \geqslant 0$, $\lambda_1, \lambda_2 \in \mathbb{R}$, **使得**

$$\begin{array}{rcl}
V_{11}\bar{w}_1 + V_{12}\bar{w}_2 + \cdots + V_{1n}\bar{w}_n - \dfrac{1}{2}\lambda_1\mu_1 - \dfrac{1}{2}\lambda_2 & = & \nu_1 \\
V_{21}\bar{w}_1 + V_{22}\bar{w}_2 + \cdots + V_{2n}\bar{w}_n - \dfrac{1}{2}\lambda_1\mu_2 - \dfrac{1}{2}\lambda_2 & = & \nu_2 \\
\cdots\cdots\cdots\cdots & \cdots & \cdots \\
V_{n1}\bar{w}_1 + V_{n2}\bar{w}_2 + \cdots + V_{nn}\bar{w}_n - \dfrac{1}{2}\lambda_1\mu_n - \dfrac{1}{2}\lambda_2 & = & \nu_n
\end{array}$$

或者

$$2V\bar{w} - \lambda_1 \mu - \lambda_2 e - \nu = 0.$$

以及

$$\bar{w}_i \geqslant 0, \qquad \nu_i \bar{w}_i = 0, \quad i = 1, \ldots, n.$$

五、Ross 的套利定价理论 (APT) 和资产定价基本定理

本讲要求：介绍 Ross 的套利定价理论 (APT)。指出 APT 方法、渐近无套利假设与线性定价法则的联系。讨论和证明有限状态情况下的资产定价基本定理。指出未定市场一般经济均衡模型与资产定价基本定理之间的关系。最后通过一个简单例子来理解资产定价基本定理怎样用来定价。

数学预备知识：线性代数。初等概率论 (随机变量的数学期望、方差、协方差等)。初等微分学（求导数，条件极值的必要条件)。

内容：

- 5.1 渐近无套利假设和 Ross 的 APT 方法。
- 5.2 多因子模型与随机折现因子。
- 5.3 有限状态情况下的资产定价基本定理。
- 5.4 从 Arrow–Debreu 证券出发来考虑资产定价基本定理。
- 5.5 资产定价基本定理的证明。
- 5.6 凸集分离定理与资产定价基本定理。
- 5.7 未定市场的一般经济均衡和资产定价第二基本定理。
- 5.8 说明资产定价基本定理的一个简单例子。
- 附录：数学预备知识 3。

“套利定价理论”是 Arbitrage Pricing Theory 的翻译，其缩写就是 APT. 我们前面用线性定价法则来为未定权益定价应该说就是一种套利定价理论，因为线性定价法则实际上是某一层次的无套利假设。但是一说起 APT, 人们往往认为这是 Ross 的“专利”。而 Ross 的 APT 则是指资本资产定价模型 (CAPM) 的一种替代物，它在形式上是把 CAPM 的单因子模型变为一个多因子模型。

5.1　Ross 于 1976 年提出的 APT ([13]) 最初也是采用类似 Sharpe 论证 CAPM 的方法，即其中蕴含效用最大化和一般经济均衡模型。但是他避开了均值－方差前沿和“市场组合”的概念，提出了类似线性定价法则的思想。由于他从随机收益出发来考虑问题，又要通过概率论的大数定律来求出均值收益，从而把问题归结为怎样用有限种证券来对无限种证券“近似”定价。最后在一些相当累赘的假定下，证得了一个逼近结果。后来 Huberman [8] 发现，为导得 APT 方法只需要“渐近无套利假设”，这就使 APT 的本质更为突出。其实，所谓渐近无套利

假设就是在线性定价法则中，要求定价线性函数连续。这一讲，我们首先将用这一观点来叙述 Huberman 的 APT 方法论证。

我们的出发点仍然是未定权益 Hilbert 空间 $\mathcal{M}$. 假定其中包含无风险证券[1]，即 $1 \in \mathcal{M}$, 并且其上有线性定价函数 $p\colon \to \mathcal{M}$ (暂时不假定它的连续性)。无风险利率 $r_f = 1/p(1) > 0$. 又假定在所涉及的未定权益市场中有无限种证券，其收益率分别为随机变量 $r_1, r_2, \ldots, r_k, \ldots \in R_1$. 同时，又假设有 K 种基本因素 (随机变量，实际应用时它们可能是各种证券的收益率，各种证券市场指数的增长率，各种宏观经济指标的增长率，以至通货膨胀率，人口出生率等等) $f_1, f_2, \ldots, f_K$, 使得下列近似等式成立：

$$r_j - r_f = \sum_{k=1}^{K} \lambda_{jk}(f_k - r_f) + \varepsilon_j, \qquad j = 1, 2, \ldots, \tag{1}$$

其中 ε_j 是一系列随机“干扰”，它们满足

$$\mathrm{Cov}[\varepsilon_i, \varepsilon_j] = 0, \qquad i \neq j, \tag{2}$$

$$\mathrm{Var}[\varepsilon_j] \leqslant \sigma^2, \qquad j = 1, 2, \ldots. \tag{3}$$

这 K 个基本因素，虽然一般都不一定是市场中的证券或证券组合，但是从抽象的观点来看，它们都要作为 $\mathcal{M}$ 的收益率 (即 R_1 的元素) 来处理。对 (1) 的两端取数学期望，可得

$$E[r_j] - r_f = \sum_{k=1}^{K} \lambda_{jk}(E[f_k] - r_f) + E[\varepsilon_j], \qquad j = 1, 2, \ldots. \tag{4}$$

如果 $\mathcal{M}$ 是一个由这 K 个基本因素所张成的有限维空间，那么 (1) 中可取所有 $\varepsilon_j = 0$. 然而，如果 $\mathcal{M}$ 是无限维空间，或者其维数远大于 K, 那么 (1)–(3) 就是一些值得论证的假设。其中假定有依赖于 j 的“误差项”，并且不同证券间的“误差项”是不相关的 (即 (2))；而所有的“误差项”的方差是一致有界的 (即 (3))。它表示所有证券的收益率都可用 K 个因素和一个方差有限、相互不相关的“误差项”来决定。不过这些都是我们问题的假设，而不是导得的结论。问题在于，如果这些“误差项”对收益的计算影响极大，那么这样的模型是无用的。而 APT 方法的关键则是指出这些“误差项”的影响会随着考虑的因素越来越多而越来越小。

---

[1] 以下的讨论其实很容易推广到不存在无风险证券的情形。其实 Ross 的经典论文中并未假定这点。

在上一假设中，(2) 和 (3) 实际上说明在模型中存在无限种收益率互不相关、而方差一致有界的证券组合。对于这样的证券组合，我们可通过对它们的再组合，使它们收益率方差 (风险) 趋向于零。事实上，对于方差有下列等式成立：

$$\operatorname{Var}\left[\frac{z_1+z_2+\cdots+z_n}{n}\right]=\frac{\operatorname{Var}[z_1]+\operatorname{Var}[z_2]+\cdots+\operatorname{Var}[z_n]}{n^2},$$

其中 $z_1, z_2, \ldots, z_n$ 为互不相关的随机变量。如果它们的方差 $\operatorname{Var}[z_i]$ 都不大于 σ^2，那么上式的右端将不大于 σ^2/n. 当 $n\to\infty$ 时，$\sigma^2/n\to 0$. 我们已经知道，如果线性定价法则成立，那么无风险利率只有一个。但是风险越来越小的证券组合，其收益率是否也趋向于无风险利率呢？这并非是显然的事。在 Huberman 的论文中，他把这一点称作“渐近无套利假设”来引入。这里，我们将首先指出，这一假设其实与要求线性定价函数连续是等价的。

定理 5.1　设 $\mathcal{M}$ 为未定权益 Hilbert 空间。$\mathcal{M}$ 中存在无风险证券，即 $1\in\mathcal{M}$. $p\colon\mathcal{M}\to\mathbb{R}$ 为 $\mathcal{M}$ 上的线性定价函数，且 $p(1)>0$, $r_f=1/p(1)$. $R_1=\{r\in\mathcal{M}\,|\,p(r)=1.\}$ 为收益率集合。那么下列两个命题等价：

1. p 为连续函数;
2. **(渐近无套利假设)**如果收益序列 $\{r_n\}\subset R_1$ 满足 $\operatorname{Var}[r_n]\to 0$, 那么 $\|r_n-r_f\|^2=E[(r_n-r_f)^2]\to 0$.

证明　假设 p 为 $\mathcal{M}$ 上的连续函数。如果收益序列 $\{r_n\}\subset R_1$ 满足 $\operatorname{Var}[r_n]\to 0$, 那么令 $\varepsilon_n=r_n-E[r_n]$, 我们有 $\|\varepsilon_n\|^2=E[\varepsilon_n^2]\to 0$. 因此，$p(\varepsilon_n)\to 0$, 即 $p(r_n-E[r_n])=p(r_n)-E[r_n]p(1)=1-E[r_n]p(1)\to 0$. 这样，我们得到 $E[r_n]\to r_f$. 另一方面，由 $\mathcal{M}$ 中的范数的三角不等式 ($\|x+y\|\leqslant\|x\|+\|y\|$), 又可得

$$\begin{aligned}\|r_n-r_f\| &= E^{1/2}[(r_n-r_f)^2]\leqslant\|r_n-E[r_n]\|+\|E[r_n]-r_f\|\\ &= \operatorname{Var}^{1/2}[r_n]+|E[r_n]-r_f|\to 0.\end{aligned}$$

因此， $E[(r_n-r_f)^2]\to 0$.

反之，令“渐近无套利假设”成立。如果 p 不连续，那么存在点列 $\{x_n\}\subset\mathcal{M}$, 满足 $x_n\to\bar{x}$ (即 $\|x_n-\bar{x}\|=E^{1/2}[(x_n-\bar{x})^2]\to 0$), 但 $p(x_n)$ 不趋向于 $p(\bar{x})$. 不妨假设存在 $\delta>0$, 使得对于所有 n，有 $|p(x_n)-p(\bar{x})|>\delta$ (否则可通过对 $\{x_n\}$ 取子列来达到这一点)。这时，可取 $r_n=(x_n-\bar{x})/(p(x_n)-p(\bar{x}))\in R_1$. 对于这个 R_1 中的点列 $\{r_n\}$, 我们有

$$\operatorname{Var}[r_n]=\operatorname{Var}\left[\frac{x_n-\bar{x}}{p(x_n)-p(\bar{x})}\right]\leqslant\frac{\operatorname{Var}[x_n-\bar{x}]}{\delta^2}\to 0.$$

从而由假设，应该有 $E[(r_n - r_f)^2] \to 0$, 即 $\|r_n - r_f\| \to 0$. 但这是不可能的。这是因为

$$\|r_n\|^2 = E[r_n^2] = E\left[\left(\frac{x_n - \bar{x}}{p(x_n) - p(\bar{x})}\right)^2\right] \leqslant \frac{\|x_n - \bar{x}\|^2}{\delta^2} \to 0,$$

由极限的唯一性，r_n 不可能再收敛于非零的 r_f. 这就反证了 p 的连续性。 □

现在我们就在定价函数 p 是连续线性函数或“渐近无套利假设”的条件下，来证明 Ross 的 APT 的主要结论。

定理 5.2 假定渐近无套利假设和 (1)–(3) 成立。那么

$$\lim_{n\to\infty} \sum_{j=1}^{n} (E[\varepsilon_j])^2 < +\infty. \tag{5}$$

证明 我们用反证法来证明。假定 (5) 不成立。那么

$$\sum_{j=1}^{\infty} (E[\varepsilon_j])^2 = +\infty.$$

不妨假定前 $K+1$ 种证券的收益率就是 $r_f, f_1, f_2, \ldots, f_K$, 即 $r_1 = f_1, \ldots, r_K = f_K$, 并且 $n > K+1$. 令

$$w_i^n = \frac{E[\varepsilon_i]}{\sum_{j=K+1}^{n} (E[\varepsilon_j])^2}, \qquad i = K+1, K+2, \ldots,$$

$$w_k^n = -\sum_{i=K+1}^{n} \lambda_{ik} w_i^n, \qquad k = 1, \ldots, K, \qquad w_0^n = -\sum_{i=1}^{n} w_i^n + 1.$$

则对于任何 n 有 $\sum_{i=0}^{n} w_i^n = 1$, 并且

$$\begin{aligned}
\sum_{i=0}^{n} w_i^n r_i &= w_0^n r_f + \sum_{k=1}^{K} w_k^n f_k + \sum_{i=K+1}^{n} w_i^n r_i \\
&= r_f + \sum_{k=1}^{K} w_k^n (f_k - r_f) + \sum_{i=K+1}^{n} w_i^n (r_i - r_f) \\
&= r_f - \sum_{k=1}^{K} \sum_{i=K+1}^{n} \lambda_{ik} w_i^n (f_k - r_f) + \sum_{i=K+1}^{n} w_i^n (r_i - r_f)
\end{aligned}$$

$$\begin{aligned} &= r_f - \sum_{i=K+1}^{n} w_i^n \left(\sum_{k=1}^{K} \lambda_{ik}(f_k - r_f) \right) + \sum_{i=K+1}^{n} w_i^n (r_i - r_f) \\ &= r_f + \sum_{i=K+1}^{n} w_i^n \varepsilon_i, \end{aligned}$$

这里的最后一步是根据 (1) 导得的。对最后一式的两端求数学期望，即得

$$\sum_{i=0}^{n} w_i^n E[r_i] = r_f + \sum_{i=K+1}^{n} w_i^n E[\varepsilon_i].$$

但是由于 w_i^n 对于 $i = K+1, K+2, \ldots,$ 的定义，由上式将导得对于所有 n 有

$$\sum_{i=0}^{n} w_i^n E[r_i] = r_f + 1, \tag{6}$$

而

$$\begin{aligned} \operatorname{Var}\left(\sum_{j=0}^{n} w_j^n r_j \right) &= \sum_{i=K+1}^{n} (w_i^n)^2 \operatorname{Var}[\varepsilon_i] \\ &\leqslant \sigma^2 \sum_{i=K+1}^{n} (w_i^n)^2 \\ &= \sigma^2 \frac{1}{\sum_{i=K+1}^{n} (E[\varepsilon_i])^2} \to 0. \end{aligned}$$

由于对任何 n, 有 $\tilde{r}_n = \sum_{j=0}^{n} w_j^n r_j \in R_1$, 由渐近无套利假设，$\sum_{i=0}^{n} w_i^n E[r_i] \to r_f$. 与 (6) 矛盾。这样，定理得证。 □

推论 在 (1)–(3) 和渐近无套利假设下，对于任何正数 $\delta > 0$, 总能找到自然数 N_δ, 使得当 $j > N_\delta$ 时，

$$(E[\varepsilon_j])^2 < \delta. \tag{7}$$

这是由 (5) 式的通项趋于零而直接得到的结论。

由此可见，(1) 的无限个等式中，除了有限个以外，都是非常好的近似等式。现在的问题是这无限个等式是否有可能被验证。如果完全要按严格的数学意义来验证，那自然是不可能的。首先现实世界中很难列举出无限种证券来，其次是近似等式要求“除去有限个以外”才成立，而这“有限个”可能是一个非常大的数目。如此等等。但是一个数

学模型是否符合实际从来都不是按严格的数学意义来看的，而是看它是否与现实充分接近 (这里又需要某种 Friedman 工具主义来作为理由，事实上，关于 APT 能否被检验的争论也是这样平息的)。在这一意义下，许多实证研究指出， APT 建议的近似等式是相当符合实际的，甚至可以要求 ε_j 与 f_k 不相关以及 $E[\varepsilon_j]=0$, 即在某种意义下，把 $\mathcal{M}$ 当作维数很小的“有限维空间”也是近似成立的。

这样一来，真正在实证分析中应用的 APT 理论其实就是一个多因子线性模型。它可以记作

$$\begin{aligned}
r_j - r_f &= \sum_{k=1}^{K} \lambda_{jk}(f_k - r_f) + \varepsilon_j, \qquad j = 1,2,\ldots, & (8)\\
\mathrm{Cov}[\varepsilon_i, \varepsilon_j] &= 0, \quad \mathrm{Cov}[f_i, f_j] = 0, \quad i \neq j, & (9)\\
\mathrm{Cov}[f_k, \varepsilon_j] &= 0, \quad k = 1,\ldots,K; \quad j = 1,2,\ldots, & (10)\\
E[\varepsilon_j] &= 0, \quad \mathrm{Var}[\varepsilon_j] \leqslant \sigma^2, \quad j = 1,2,\ldots. & (11)\\
E[r_j] - r_f &= \sum_{k=1}^{K} \lambda_{jk} E([f_k] - r_f), \qquad j = 1,2,\ldots. & (12)
\end{aligned}$$

这里条件 $\mathrm{Cov}[f_i, f_j]=0$ $(i \neq j)$ 又是新增的。这一条件并不很苛刻。一般总认为所考虑的因子之间是不相关的。即使它们不满足这一条件，也可通过因子组的适当的线性变换 (即用因子的线性组合代替原来的因子) 来使其满足。在这种情况下，对 (8) 的两端乘上 $f_l - E[f_l]$ 并取数学期望，我们得到

$$E[(r_j - r_f)(f_l - E[f_l])] = \sum_{k=1}^{K} \lambda_{jk} E[(f_k - r_f)(f_l - E[f_l])] + E[\varepsilon_j(f_l - E[f_l])].$$

因为 r_f 和 $E[f_l]$ 是常数，我们有

$$E[r_f(f_l - E[f_l])] = E[E[f_l](f_l - E[f_l])] = 0.$$

从而

$$\begin{aligned}
E[(r_j - r_f)(f_l - E[f_l])] &= E[(r_j - E[r_j])(f_l - E[f_l])] - E[r_f(f_l - E[f_l])]\\
&= \mathrm{Cov}[r_j, f_l].
\end{aligned}$$

同理可证，

$$E[(f_k - r_f)(f_l - E[f_l])] = \mathrm{Cov}[f_l, f_k], \; E[\varepsilon_j(f_l - E[f_l])] = \mathrm{Cov}[\varepsilon_j, f_l] = 0.$$

由此可得，

$$\lambda_{jk} = \beta_{r_j,f_k} = \frac{\mathrm{Cov}[r_j, f_k]}{\mathrm{Var}[f_k]}, \quad k = 1, 2, \ldots, K;\ j = 1, 2, \ldots.$$

尤其是，当 $K = 1$ 时，令 $r_m = f_1$, $\beta_j = \lambda_{1j}$, 我们有

$$r_j - r_f = \beta_{r_j,r_m}(r_m - r_f) + \varepsilon_j, \qquad j = 1, 2, \ldots.$$

$$E[r_j] - r_f = \beta_{r_j,r_m}(E[r_m] - r_f), \qquad \beta_{r_j,r_m} = \mathrm{Cov}[r_j, r_m]/\mathrm{Var}[r_m].$$

这似乎又回到资本资产定价模型 (CAPM)。但是就如 APT 并非是简单的多因子线性模型一样， CAPM 也不说是 APT 的形式推论。我们仔细考察上面的 APT 的推导过程，就会发现，这个形式上的 CAPM 与以前讨论的 CAPM 在含义上是很不一样的。

5.2 我们曾经在前面讨论过 CAPM 与随机折现因子理论 (线性定价法则) 之间的等价性。在上一节的最后，我们指出了 APT 在形式上可导出类似于 CAPM 的表达式。这就可提出这样的一个问题：如果有一个很一般的 APT 形式的表达式成立，那么它与随机折现因子理论有什么关系？正如我们在讨论 CAPM 时那样，我们应该也有一个随机折现因子 m 与 APT 中的因子之间的关系。为此，我们指出下列一般多因子模型与随机折现因子之间的关系。

定理 5.3 设 $\mathcal{M}$ 为未定权益 Hilbert 空间。 $p\colon \mathcal{M} \to \mathbb{R}$ 为其连续线性定价函数， $m \in \mathcal{M}$ 为其相应的随机折现因子，即对于任何 $x \in \mathcal{M}$, 有 $p(x) = E[mx]$. $r_f = 1/E[m] > 0$ 为其相应的无风险利率。 $R_1 = \{r \in \mathcal{M} \,|\, p(r) = E[mr] = 1.\}$ 为其相应收益超平面。 $g_1, g_2, \ldots, g_K \in \mathcal{M}$, 为方差非零的随机变量，且 $E[g_1] = E[g_2] = \cdots = E[g_K] = 0$. 那么存在常数 $a_1, \ldots, a_K \in \mathbb{R}$ 使得 $m = E[m] + a_1 g_1 + \cdots + a_K g_K$ 当且仅当对于任何 $r \in R_1$, 有

$$E[r] - r_f = \beta_{1r}\mu_1 + \beta_{2r}\mu_2 + \cdots + \beta_{Kr}\mu_K,$$

其中

$$\beta_{ir} = \mathrm{Cov}[g_i, r]/\mathrm{Var}[g_i], \quad \mu_i = -a_i r_f \mathrm{Var}[g_i], \quad i = 1, 2, \ldots, K. \tag{13}$$

证明 事实上，如果 $m = E[m] + a_1 g_1 + \cdots + a_K g_K$, 那么对于任何收益 $r \in R_1$, 有 $E[mr] = E[m]E[r] + \mathrm{Cov}[m, r] = 1$. 因此，由 $r_f = 1/E[m]$, 可得

$$\begin{aligned} E[r] &= r_f - r_f \mathrm{Cov}[m, r] = r_f - r_f \mathrm{Cov}[a_1 g_1 + \cdots + a_K g_K, r] \\ &= r_f + \beta_{1r}\mu_1 + \cdots + \beta_{Kr}\mu_K, \end{aligned}$$

其中 β_{ir}, μ_i 如 (13) 所定义。反之，如果上式成立，我们也可取 $a_i = -\mu_i/r_f \mathrm{Var}[g_i]$, $i = 1, \ldots, K$, 使得 $m = E[m] + a_1 g_1 + \cdots + a_K g_K$ 成立。□

定理 5.3 说明，如果未定权益的期望收益可用一个多因子模型来刻划，那么这等价于其随机折现因子可用一个多因子线性组合来刻划。回顾上一节最后所述实际使用的 APT 模型 (8)–(12)，它其实也可以归结为这种多因子模型。事实上，我们在上面已经导得

$$E[r_j] - r_f = \sum_{k=1}^{K} \beta_{r_j f_k} E([f_k] - r_f), \qquad j = 1, 2, \ldots.$$

这里的 f_k 与上面的 g_k 的区别在于它不一定满足 $E[f_k] = 0$. 同时，这个表达式并未要求对于所有收益率 $r_j \in R_1$ 满足。但是如果满足该模型的 r_j 已经构成 R_1 中的一个“基”，即每个收益率 $r \in R_1$ 都可表示为 r_j 的仿射组合或者它们的极限时，我们也可认为上述表达式对于所有 $r \in R_1$ 满足。这时，我们令 $g_k = f_k - E[f_k]$, 那么 $\mathrm{Var}[g_k] = \mathrm{Var}[f_k]$, 并且对于任何 r 有 $\mathrm{Cov}[r, g_k] = \mathrm{Cov}[r, f_k]$. 由此可得，

$$E[r] - r_f = \sum_{k=1}^{K} \beta_{kr}(E[f_k] - r_f), \quad \beta_{ir} = \mathrm{Cov}[g_i, r]/\mathrm{Var}[g_i].$$

与上面的表达式相比较，即得

$$\mu_i = -a_i r_f \mathrm{Var}[g_i] = E[f_i] - r_f, \quad i = 1, 2, \ldots, K.$$

因此，上述“APT 多因子模型”实际上相当于定义了随机折现因子

$$m = \frac{1}{r_f} + \sum_{i=1}^{K} \frac{(r_f - E[f_i])(f_i - E[f_i])}{r_f \mathrm{Var}[f_i]}.$$

再回到基本定价公式 $p(x) = E[mx]$, 我们立即看到

$$p(x) = E[m]E[x] + \mathrm{Cov}[x, m] = \frac{E[x]}{r_f} + \sum_{i=1}^{K} (r_f - E[f_i])\beta_{xf_i},$$

$$\beta_{xf_i} = \frac{\mathrm{Cov}[x, f_i]}{\mathrm{Var}[f_i]}, \quad i = 1, \ldots, K.$$

即这 K 个因子刻划了未定权益的风险价值，并且其大小是相应的 β 的线性组合。这个结果甚至对于 $K = 1$ 时 (相当于 CAPM 的情形) 也很有意义。

5.3 Ross 在提出他的 APT 理论以后，进一步又结合 Black–Scholes 期权定价理论的发展，于 1978 年在 [14] 中提出一条很一般的定理。这条定理后来被人们称为资产定价基本定理。它指出完整的无套利假设等价于正线性定价法则。当时，如我们前面所叙述的框架和线性定价法则尚不为人们所理解和熟悉。因此，这条资产定价基本定理对金融经济学框架的形成，实际上起了决定性的作用。正如，Ross 在他的论文的前言中所说：“在一个没有未被开发的套利机会的资产市场中，存在一个线性估值算子，它可以毫不含糊地以完善的市场替代来为收益流定价，或者对通过市场组合界定的现金流来界定其值。用不到进一步假定，只要预计的收益可以通过购买一个市场资产组合的确定的跨时规划来复制 (或界定)，这是可能的。这些结果已被证明，并且被用来简化和统一许多金融经济学中的论述，其中包括项目估值，Modigliani-Miller 理论，远期定价，封闭式互助基金悖论以及有效市场理论。”

资产定价基本定理在数学上涉及一条深刻的数学定理：凸集分离定理。为便于理解，我们将先把我们的模型简化：假定未定权益在未来只可能取有限个值；也就是说，未来只可能发生有限的 S 种状态，而每个随机变量也就只能取 S 个值。这样，在不指定这 S 种状态所发生的概率时，一个随机变量也就与一个 S 维向量等同起来。这种假定未来只有有限种状态但又不指定它们所发生的概率的讨论方式在历史上是 Arrow 和 Debreu 在他们的经典的一般经济均衡研究中首先用的。在这样的模型中，不同时间、不同状态下的同一种商品被看作不同的商品。由此 Arrow–Debreu 就可通过扩大商品空间维数的方法把确定性的市场的一般经济均衡理论推广到带不确定性的市场情形。这样的推广虽然富有启发性，但并未抓住金融市场的不确定性本质。事实上，用后人的话来说，他们的这种做法相当于对每一种不确定状态都引进了一种有单位价值的证券，从而使每一种未定权益都可表示为这种证券的线性组合。这样的市场一定是完全市场，它意味着市场中没有规避不了的风险。而对金融经济学来说，更重要的是不完全市场，即存在不能被基本证券的线性组合所复制的未定权益的市场。尽管如此，Arrow–Debreu 的先驱研究为金融经济学文献留下了两个重要术语。一是所谓“**Arrow–Debreu 证券**”。在数学上，它们就是一个 S 维向量空间中的单位向量。它们代表只在一种状态下取 1，其他状态下取 0 的未定权益。另一个是所谓“**未定市场** (contingent market)” (见例如 [12])，或“未定权益市场 (contingent claim market)” (见例如 [2]，不过我们不想使用这后一个术语，以免与我们的“未定权益空间”混淆)。这两个术语决不能从它们的字面意思去理解，它们其实都是指一个未来只有有限种状态的二期证券市场模型。换句话说，未定市场就是未定权益用 (有限) S 维向量来刻划的市场，而通常又不指定这 S 种状态所发生的概率。

对未来可能发生的状态不指定概率的讨论在经济学上起源于芝加哥经济学派的先驱者 F. H. Knight (1885-1972), 他曾经明确地把带概率分布的不确定性与无法指定概率分布的不确定性区分开来，并且把前者称为“风险 (risk)”，后者才称为“不确定性 (uncertainty)”。虽然后人没有严格按 Knight 的主张来做，但 Arrow–Debreu 显然是受到了 Knight 的影响。这种不带概率分布的不确定性一般只能对于有限状态的情形来讨论，而这种讨论当然对人们认识金融现实很有好处。事实上，在许多金融实际问题上，对未来事件发生的可能性往往是各有各的估计，或者说对所涉及的随机变量各有各的概率分布估计。这种各自自行定义的概率称为**主观概率**。如果我们假定未来只有 S 种可能性，那么每一种 (主观) 概率测度可以对每种状态发生的概率估计有所不同，但其共同点是这 S 种状态发生的概率都不为零，并且它们的总和是 1. 所有这样的在 S 种状态上定义的概率测度都称为这个 S 种状态的概率空间上的**等价概率测度**。假定未来只有 S 种状态使得数学上只需要古典概率论，而避免了许多数学困难。但是它仍然能得到许多金融经济学上的有意义的结论。这些结论的相当部分都不一定能推广到连续变化的随机变量情形。对于从事实证分析的金融经济学家来说，或许这种针对有限状态的结论已经足够，但是对于从事理论研究的金融经济学家和数学家来说，就有许多细腻的数学问题需要仔细推敲。

我们下面的讨论将是在这样的框架中证明资产定价基本定理：完整的无套利假设等价于定价函数一定是正线性函数，并且它可以用一种等价概率测度来表示 (我们在上节中对 $S=2$ 的最简单的模型已经得到过这种等价概率测度)。为此，我们先来进一步明确“未定权益空间 $\mathcal{M}$”. 在目前的情形下，我们可假定 $\mathcal{M}$ 就是 S 维向量空间 $\mathbb{R}^S$. 在这个 $\mathcal{M}=\mathbb{R}^S$ 中，有 K 种基本证券，它们的未来的价格都有 S 种可能。我们仍以 x_i 来表示第 i 种证券未来的价格，它被假定为一个 S 维行向量 $x_i=(x_i^1,x_i^2,\ldots,x_i^S)\in\mathbb{R}^S$, 其中上标 $s=1,2,\ldots,S$ 仅仅是编号，而不是幂，它们代表第 s 种状态时的证券价格。也可以说，x_i 是一个 $1\times S$ 阶矩阵。这些证券的未来价格的全体仍然可以表示为 $X=(x_1,\ldots,x_K)^T$, 但是它实际上是一个 $K\times S$ 阶矩阵。X 常被称为**支付矩阵** (payoff matrix)。组合 $\theta=(\theta_1,\ldots,\theta_K)$ 是 $1\times K$ 阶矩阵，即仍然是一个 K 维行向量，但是 θ 与 X 的乘积

$$\theta\cdot X=\theta_1x_1+\cdots+\theta_Kx_K\in\mathbb{R}^S$$

将是一个 $1\times S$ 阶矩阵 (行向量)，表示这一证券在未来 S 种状态下可能有的值。此外，与前面一样，定价问题仍然归结为确定一个定价函数 p, 但是 $p\colon\mathbb{R}^S\to\mathbb{R}$ 将是一个 S 维空间上的函数，并且认为它在 K 种基本证券上的值 $p(x_i)$, $i=1,2,\ldots,K$, 已经确定。

这样，我们的出发点是：未定权益空间 $\mathcal{M}=\mathbb{R}^S$ 为 S 维空间，其

上有 K 种基本证券，它们的未来价格由支付矩阵 $X=(x_1,\ldots,x_K)^T$ 来刻划，这里 $x_i=(x_i^1,x_i^2,\ldots,x_i^S)\in\mathbb{R}^S$, 而定价函数 p 已经在这 K 种证券上有定价：$p(X)=(p(x_1),p(x_2),\ldots,p(x_S))^T$. 我们记这样的证券市场为 $(\mathbb{R}^S,X,p(X))$. 我们对这个市场提出下列无套利假设：

$(\mathbb{R}^S,X,p(X))$ **上的无套利假设**：

1. 线性定价法则成立，即 p 是一个 $\mathcal{M}=\mathbb{R}^S$ 上的线性函数。因此，对于任何组合 $\theta=(\theta_1,\ldots,\theta_K)$, 其当前价值为
$$p(\theta\cdot X)=p(\theta_1x_1+\cdots+\theta_Kx_K)=\theta_1p(x_1)+\cdots+\theta_Kp(x_K).$$
2. p 对于可交易未定权益 $y=\theta\cdot X$ 是正线性函数。即如果 $y=\theta\cdot X\in\mathbb{R}^S$ 满足 $y>0$, 那么 $p(y)>0$. 这里 $y>0$ 表示 y 的分量非负，并且至少有一个为正。

这一无套利假设首先是对支付矩阵 X 和定价向量 $p(X)$ 作出了限制；即对 K 种基本证券的当前价格作出限制。例如，第 i 种证券的未来价格 $x_i>0$, 那么它的当前价格 $p(x_i)$ 必须也大于零。但仅仅这样还不够。还要考虑各种基本证券当前价格之间的协调。否则可能根本不存在满足上述无套利假设的定价函数。其实，我们在上一讲中已经在一个较简单的模型中讨论过这个问题。在那里，我们只有一种风险证券，即股票，其中假定当前股价 S_0 已知，而未来股价或是变成 aS_0, 或是变成 bS_0. 另一种证券是无风险证券，它的未来价格和当前价格始终为 1. 无套利假设在那里可导得 $a<1<b$. 现在的问题是当 $K>2$ 时，各种证券的当前价格与未来价格应该满足什么条件，才与无套利假设不矛盾。这一问题就是 Ross 在 1978 年所要解决的，它就是我们要讨论的**资产定价基本定理**。

定理 5.4 $(\mathbb{R}^S,X,p(X))$ 上的无套利假设成立的充分必要条件为存在 S 个正数 $\lambda_1,\ldots,\lambda_S>0$, 使得
$$p(x_i)=\lambda_1x_i^1+\cdots+\lambda_Sx_i^S,\qquad i=1,\ldots,K.\tag{14}$$

这条定理说明 $(\mathbb{R}^S,X,p(X))$ 上的无套利假设意味着证券的当前价格是证券的未来各种可能的价格的正线性组合。尤其是当第 1 种证券是无风险证券，其未来价格为 $x_1=\mathbf{1}=(1,1,\ldots,1)$ 时，如果 $p(x_1)=p(\mathbf{1})=1/r^f>0$, 那么 $1/r_f=\sum_{s=1}^S\lambda_s$. 令 $p_s=\lambda_sr_f$, $s=1,\ldots,S$. 则 $\sum_{s=1}^S p_s=1$. 从而，p_s 可以看作第 s 种状态在未来发生的概率。这样，(14) 式又变为
$$p(x_i)r_f=p_1x_i^1+\cdots+p_Sx_i^S,\qquad i=0,1,\ldots,K.$$

上式右端是证券未来价值对于概率 $p_1, \ldots, p_S$ 的均值 (数学期望)，左端是证券的当前价值的无风险获利。于是上述定理又可表达为

定理 5.4’ 如果第 1 种证券是无风险证券，其未来价格为 $x_1 = \mathbf{1} = (1, 1, \ldots, 1)$，其当前价格为 $p(x_1) = p(\mathbf{1}) = 1/r^f > 0$. 那么 $(\mathbb{R}^S, X, p(X))$ 上的无套利假设等价于存在某种等价概率测度，使得每一种未来价格为 x_i 的基本证券的收益率 $(x_i/p(x_i))$ 的数学期望都等于无风险利率 r_f.

这条定理的充分性很好证明。事实上，设 (14) 式成立。如果 $y = \theta \cdot X > 0$，那么

$$p(\theta \cdot X) = \sum_{i=0}^{K} \theta_i p(x_i) = \sum_{i=0}^{K} \sum_{s=1}^{S} \theta_i \lambda_s x_i^s$$

$$= \sum_{s=1}^{S} \lambda_s \left(\sum_{i=0}^{K} \theta_i x_i^s \right).$$

由于所有 $\lambda_s > 0$，故由右端括号中的项都非负、且至少有一项为正，可导得左端为正，即上述无套利假设成立。

但是必要性的证明在数学上涉及一些比较深刻的结果，其中必须用到凸集分离定理或其等价结果。对此，我们需要一定的准备。这里我们先来论述一些其他的问题，以便于理解。

5.4 我们在上面曾经提到 Arrow–Debreu 证券的概念。它们的未来价格就是现在涉及的未定权益空间 $\mathbb{R}^S$ 中的单位向量。从数学上来看，如果基本证券就是 Arrow–Debreu 证券全体，那么定价问题的解是一目了然的。也就是说，一个定价函数如果已经对于每一单位向量都有确定的 (正) 值，那么它对任何向量的取值也就完全确定。这是因为任何向量都可以用这些单位向量来线性表示。而资产定价基本定理实际上是指出，每一种符合无套利假设的定价法则都必须是对所有 Arrow–Debreu 证券取正值的定价法则。换句话说，对每一种 Arrow–Debreu 证券都取正值的定价函数一定满足无套利假设；而满足无套利假设的定价函数也一定使 Arrow–Debreu 证券取正值。用这样的方式来表达资产定价基本定理使得这条定理变得极为自然，甚至给人一种错觉，似乎资产定价基本定理是不证自明的，用不到什么凸集分离定理之类。其实这是因为我们假定基本证券就是 Arrow–Debreu 证券所引起的。一般情形没有那么简单。

但是从向量空间的观点来看，一般的资产定价基本定理无非就是要搞清基本证券组和 Arrow–Debreu 证券组之间的关系。这是两个向量

组。后一组是向量空间的单位向量组，任何向量都可以用它们来线性表示，因此，基本证券组中的每一个向量也都可用它们来线性表示。反之则不一定，因为基本证券组不一定构成向量空间的基。如果基本证券组构成向量空间的基，那么 Arrow–Debreu 证券组中的每一个向量也都能用基本证券组中的向量来线性表示。在这种情形下，资产定价基本定理的证明就非常简单。其证明的大意可叙述如下 (参看后面的叙述)：由于所有 Arrow–Debreu 证券都等价于某种基本证券的组合，而对于 Arrow–Debreu 证券集来说，无套利假设就等价于定价函数对所有 Arrow–Debreu 证券取正值。因此，资产定价基本定理成立。这里还不难看出，定理 5.4 中的 λ_i 就是第 i 个 Arrow–Debreu 证券的价格。因此，它们还常被称为**状态价格**。由于基本证券集构成向量空间基的情形处理起来特别简单，这就有必要再次强调一个特别的金融学概念。它就是**完全市场**的概念。在这种情形下，每一个 S 维向量都可以用基本证券集中的向量来线性表示。利用向量空间理论中的已知结论，可以断定，这时必须有 $K \geqslant S$, 即证券的种类数要超过未来状态的种类数。在现实世界中，这样的情形是不可能真正发生的。因此，完全市场只是一种理想情形。在这一理想情形下，上述数学条件翻译成金融语言，就变成：每一种未定权益 (资产、期权、衍生证券) 都可用基本证券 (标的资产) 来复制 (为基本证券的组合)。

那么不完全市场的资产定价基本定理能否通过 Arrow–Debreu 证券来论证呢？这当然可以进行尝试。不过稍加分析就会发现，由于基本证券向量不能张成全空间，必要性的证明中总会有一些麻烦存在。利用 Arrow–Debreu 证券来证明资产定价基本定理其实是可行的。事实上，我们只需要在基本证券集中再加上若干 Arrow–Debreu 证券，使扩大后的基本证券集形成完全市场即可。我们知道一组向量如果不构成向量空间的基，那么总能加上若干个与它们线性无关的向量，使扩大后的向量组构成基。而当市场不完全时，我们总能找到一个 Arrow–Debreu 证券，使得它的未来价格作为向量与基本证券集中的相应向量全体线性无关。如果加一个不够，那么还可以再加，一直加到市场完全为止。这就是说，任何基本证券集，总能通过加入若干 Arrow–Debreu 证券，使得市场完全化。问题在于如何为这些新加入的 Arrow–Debreu 证券定价，使得无套利假设仍然满足。毫无疑问，这些证券的当前价格必须是正的；否则无套利假设已经被破坏。但是是否可以取任何正数？这就不一定。例如，在我们已经见过的最简单的二状态情形中，如果基本证券集中就只有一个无风险证券，它的当前价格是 1, 而未来价格也始终是 1, 那么正如我们以前已经看到的，再加入一种 Arrow–Debreu 证券 (即其两种未来价格分别是 0 和 1) 时，该证券的价格只能在 0 和 1 之间。否则就会违背无套利假设。尽管如此，可以指出，对于任何不完全的基本证券集，我们总可以对它们加入若干适当定价的 Arrow–Debreu

证券，使得市场完全化。为证明这一点 (见下一节)，需用应用一些从数学上来说与凸集分离定理本质上是等价的技巧。但是这样的一条思路似乎有比较明确的经济意义。

5.5 现在我们利用 Arrow–Debreu 证券来证明资产定价基本定理。这里我们只需证明必要性。以下叙述中的无套利假设都是指 $(\mathbb{R}^S, X, p(X))$ 上的无套利假设。

1. 完全市场情形：这时基本证券集的未来价格 $\{x_i\}$ 构成 S 维向量空间的基。不妨假设基本证券的个数为 $K = S$, 并且 $x_1, x_2, \ldots, x_S$ 构成向量空间的基，其他的向量都是它们的 (唯一) 线性组合。尤其是所有 Arrow–Debreu 证券

$$e_i = (0, \ldots, 0, \underbrace{1}_{\text{第 } i \text{ 个}}, 0, \ldots, 0), \quad i = 1, 2, \ldots, S;$$

都是它们的 (唯一) 线性组合，即

$$e_i = a_i^1 x_1 + a_i^2 x_2 + \cdots + a_i^S x_S, \qquad i = 1, 2, \ldots, S. \tag{15}$$

同时，显然也有

$$x_i = x_i^1 e_1 + x_i^2 e_2 + \cdots + x_i^S e_S, \qquad i = 1, 2, \ldots, S. \tag{16}$$

根据无套利假设，由 (15) 可知，e_i 的当前价格 $p(e_i)$ 可由 $\{x_j\}$ 的当前价格 $\{p(x_j)\}$ 来确定，即

$$p(e_i) = \lambda_i = a_i^1 p(x_1) + a_i^2 p(x_2) + \cdots + a_i^S p(x_S), \qquad i = 1, 2, \ldots, S.$$

并且再由无套利假设可得，所有 $\lambda_i > 0$. 另一方面，由 (16) 可知，

$$p(x_i) = x_i^1 p(e_1) + x_i^2 p(e_2) + \cdots + x_i^S p(e_S) = \lambda_1 x_i^1 + \lambda_2 x_i^2 + \cdots + \lambda_S x_i^S, \ \ i = 1, 2, \ldots, S.$$

即定理得证。

2. 不完全市场情形：这时基本证券集的未来价格 $\{x_i\}$ 不构成 S 维向量空间的基。不妨假设基本证券的个数为 $N < S$, 并且 $x_1, x_2, \ldots, x_N$ 线性无关。从而，由 $\{x_i\}$ 的当前价格 $\{p(x_i)\}$ 只能对这 N 个向量所张成的 N 维空间 X_N 所代表的证券 (唯一) 定价。而对于 X_N 以外的证券无法定价。这也就是说，对于任何 $y \in X_N$, 我们定义了一个线性定价函数 $p_N \colon X_N \to \mathbb{R}$，使得对于任何 $y = \theta_1 x_1 + \cdots + \theta_N x_N$, 有

$$p_N(y) = \theta_1 p(x_1) + \cdots + \theta_N p(x_N).$$

并且由无套利假设，还可得到：如果 $y > 0$, 则 $p_N(y) > 0$. (不过，这样的 y 也有可能在 X_N 中找不到。)

由于 X_N 仅仅是 $N(< S)$ 维空间，那么至少有一个 Arrow–Debreu 证券不在其中。不妨设它是 e_j, 它的定价 $e_j^0 = \alpha$ 由无套利假设必须为正值，即 $\alpha > 0$. 定义 $X_{N+1} = \{y + ae_j \,|\, y \in X_N,\ a \in \mathbb{R}\}$. 这是个 $N+1$ 维空间，其上的定价函数 p_{N+1} 可定义为

$$\forall y \in X_N,\ a \in \mathbb{R}, \quad p_{N+1}(y + ae_j) = p_N(y) + a\alpha.$$

这个函数显然是线性函数，但对于任意取的正数 $\alpha > 0$, 不一定能保证它满足无套利假设，即不能保证 $y + ae_j > 0$ 时，有 $p_{N+1}(y + ae_j) > 0$. 我们指出，当 α 在某个区间中取值时，这一要求一定能满足 (以下的讨论在最简单的只有一种无风险证券和一种风险证券的情形下，相当于指出，已知风险证券未来的两种可能 $[S = 2]$ 的价值，那么它当前的定价只要在这两个价值之间，就满足无套利假设)。

事实上，对于任何 $y \in X_N$, 我们有

$$p_N(y) = \begin{cases} > 0, & y > 0, \\ \text{任意}, & y \not> 0. \end{cases}$$

这是一个条件不等式。为使它表示为“无条件不等式”。我们引进一个定义在整个 S 维空间中的广义实值函数 $V\colon \mathbb{R}^S \to \mathbb{R}$ 如下：

$$\forall x \in \mathbb{R}^S, \quad V(x) = \begin{cases} 0, & x > 0, \\ -\infty & x \not> 0. \end{cases}$$

那么上式也可表示为

$$\forall y \in X_N, y \neq 0, \quad p_N(y) > V(y).$$

不难验证，函数 V 满足

$$\forall x, y \in \mathbb{R}^S, \qquad V(x + y) \geqslant V(x) + V(y),$$

$$\forall \lambda \geqslant 0, \qquad V(\lambda x) = \lambda V(x).$$

前一性质称为 V 的超可加性，后一性质称为 V 的正齐次性。从而对于任何 $y, z \in X_N$, 和 $x \in \mathbb{R}^S$, 有

$$p_N(y) - p_N(z) = p_N(y - z) \geqslant V(y - z) \geqslant V(y + x) + V(-z - x),$$

即

$$-V(-z - x) - p_N(z) \geqslant V(y + x) - p_N(y).$$

取 $x = e_j$, 并在左端对 z 求下确界，而在右端对 y 求上确界，则可得

$$c = \inf_{z \in X_N} (-V(-z - e_j)) - p_N(z)) \geqslant \sup_{y \in X_N} (V(y + e_j) - p_N(y)) = b,$$

可以看出，b 是下列问题的值：

$$\begin{cases} \max(-p_N(y)) \\ y \in X_N \\ y^j \geqslant -1,\ y^i \geqslant 0,\ i \neq j. \end{cases}$$

而 c 是下列问题的值：

$$\begin{cases} \min(-p_N(z)) \\ z \in X_N \\ z^j \leqslant -1,\ z^i \leqslant 0,\ i \neq j. \end{cases}$$

显然 $c \geqslant b \geqslant 0$. 同时，必须有 $c > b$. 这是因为不难看出上述两个问题都是有解的；如果 $c = b$, 那么就存在 $\bar{y}, \bar{z} \in X_N$, 满足

$$\bar{y}^j \geqslant -1,\ \bar{y}^i \geqslant 0,\ i \neq j; \quad \bar{z}^j \leqslant -1,\ \bar{z}^i \leqslant 0,\ i \neq j.$$

以及 $p_N(\bar{y}) = p_N(\bar{z}) = -b = -c$. 但是由上式可知，

$$\bar{y}^i - \bar{z}^i \geqslant 0, \quad i = 1, 2, \ldots, S; \quad \bar{y} \neq \bar{z}.$$

这仅当 $\bar{y} - \bar{z} > 0$ 时才有可能。而 $p_N(y - z) = 0$. 这与无套利假设矛盾。因此，$c > b$.

这样就形成一个数值区间 (b, c) 我们取 $\alpha \in (b, c)$, 则 $\alpha > 0$. 另一方面，对于任何 $y \in X_N$, $a \neq 0$, 由 α 的定义，我们有

$$-V\left(-\frac{y}{a} - e_j\right) > p_{N+1}\left(\frac{y}{a} + e_j\right) = p_N\left(\frac{y}{a}\right) + \alpha > V\left(\frac{y}{a} + e_j\right),$$

注意到 V 的正齐次性，可以看出，不管 a 是正的，还是负的，我们都有

$$p_{N+1}(y + ae_j) > V(y + ae_j).$$

尤其是当 $y + ae_j > 0$ 时，$p_{N+1}(y + ae_j) > 0$. 这说明对 Arrow–Debreu 证券 e_j 的定价 α 仍能使新的基本证券集满足无套利假设。

最后，如果新的基本证券集仍然使市场不完全，我们还可以继续用同样的方法加入 Arrow–Debreu 证券。一直到得到一个保持原有定价、并形成完全市场的基本证券集为止。由此归结为上述情形。定理得证。
□

5.6 资产定价基本定理的常规证明是利用一条数学定理：凸集分离定理。利用凸集分离定理的证明完全变成数学讨论。其过程很难再与经济意义相联系。但是凸集是个在经济学中非常有用的概念，我们会在许多其他地方用到它。因此，我们对此也应该有所了解。

现在我们用凸集分离定理来证明资产定价基本定理的必要性：

在 $\mathbb{R}^{S+1}$ (这是当前状态的价值与未来的 S 种状态的价值所形成的空间) 中考虑集合

$$M = \left\{ z \in \mathbb{R}^{S+1} \,\middle|\, z = (z_0, z_1, \ldots, z_S) > 0, \sum_{s=0}^{S} z_s = 1 \right\},$$

(这是当前价值为零的 Arrow–Debreu 证券与未来价值为零、当前价值为 1 的证券全体所形成的凸集。这一集合中的证券是当前价值与未来价值都非负的非零向量。) 和集合

$$L = \{ y \in \mathbb{R}^{S+1} \,|\, y = (-\theta \cdot x^0, \theta \cdot x^1, \ldots, \theta \cdot x^S),\ \theta \in \mathbb{R}^{K+1} \}$$

这里 $x^s = (x_1^s, \ldots, x_K^s)^T$, $s = 1, \ldots, S$, $x^0 = (p(x_1), \ldots, p(x_K))^T$. ($L$ 是组合 θ 的加上负号以后的当前价值与 S 种未来价值所形成的向量全体)。那么由无套利假设，它们是两个不相交的闭凸集 (如果相交，那么意味着有一种非零组合的当前价值非正，未来价值非负)，并且不难验证 $M - L$ 是闭集。从而存在 $\mu = (\mu_0, \mu_1, \ldots, \mu_S) \in \mathbb{R}^{S+1}$ 使得

$$\forall y \in L,\ \forall z \in M, \qquad (\mu, z) > (\mu, y);$$

但由于 L 实际上是 $\mathbb{R}^{S+1}$ 中的子空间，上式的右端只可能为零，否则它可能无限增大，以至不等式不可能成立。另一方面，由于 z 可取只有一个分量为一、其他分量为零的单位向量，μ 的每一分量都必须为正，否则上述严格不等式不可能成立。因此，

$$\forall \theta \in \mathbb{R}^K, \qquad (\theta, -\mu_0 x^0 + \mu_1 x^1 + \cdots + \mu_S x^S) = 0,$$

以至

$$-\mu_0 x^0 + \mu_1 x^1 + \cdots \mu_S x^S = 0.$$

令 $\lambda_s = \mu_s/\mu_0$, 并注意到 $x^0 = (p(x_1), \ldots, p(x_K))^T$, 即得 (14) 式。 □

从定理的证明中我们可以看出，由此确定的等价概率测度一般并不唯一。事实上，它是由包含 L 的线性子空间来决定的。如果 L 本身不是 S 维子空间，那么通过 L 的满足定理条件的子空间就有无限多个，

从而也可有无限多个等价概率测度符合定理要求。在这种情况下，我们无法利用这样的等价概率测度来对定价问题作出肯定的回答。或者也可从对 $\lambda_1, \lambda_2, \ldots, \lambda_S$ 求解方程 (15) 来看。这是一组线性方程组，它不一定有解，有解也不一定有唯一解。在有多个不同的解时就一定有无限多个解，从而就有无限多个相应的等价概率测度。另一方面，(15) 作为方程无解时说明无套利假设一定不满足。这意味着，**如果** (**基本**) **证券定价不合理，就会存在套利机会**。

如果 (15) 成立，那么 L 中的向量的第一个分量完全由后面 S 个分量所决定。因此，向量空间 L 的维数也完全由其后面的 S 个分量所决定。因此，L 是 S 维子空间的充分必要条件为

$$\mathbb{R}^S = \{\, z \in \mathbb{R}^S \,|\, z = (\theta \cdot x^1, \ldots, \theta \cdot x^S),\ \theta \in \mathbb{R}^{K+1} \,\}.$$

这个条件仅当 $K+1 \geqslant S$ 时才有可能满足 (但并非一定满足)，这时 (15) 的方程个数不比 λ_s 的个数少。回顾前面提到的概念，所谓衍生证券或未定权益就是 $\mathbb{R}^S$ 中的向量 (从金融市场的角度来看，它们应该是非负向量，但这并不影响以下的数学讨论)。于是上面的等式将意味着：**任何未定权益都可通过某种证券组合来实现**。这样的证券市场就是前面已经提到的**完全市场**。从而上面的讨论实际上已经指出，**在完全市场中，每种未定权益都可唯一定价**。这是因为这时的等价概率测度是唯一的。当 $K < S$ 时，市场一定是不完全的。它可以理解为没有足够多的证券种类来应付各种可能发生的情况。对于现实世界来说，市场不可能是绝对完全的。因此，完全市场的概念只是一个理想概念。

上面的讨论都是对于有限维未定权益空间情形进行的。而 Ross 在他 1978 年的经典论文 [14] 中却是为适应期权定价理论的需要一开始就从无限维空间出发，并且也应用了凸集分离定理。不过他的数学模型和证明似乎不能令人满意。后来才逐渐发现，无限维空间上的资产定价基本定理要比有限维情形下的定理复杂得多。一般来说，无套利假设不再与连续正线性定价函数等价。其中有许多很细腻的数学技巧问题需要解决。有兴趣的读者可参看例如最近发表的 Clark 的论文 [2]。另外一些有关的问题请参看后面的说明。

至于在未定权益 Hilbert 空间的框架中，我们容易指出，当随机折现因子 m 是严格正随机变量 (即它大于零的概率为 1) 时，由 $x > 0$ 可导出 $p(x) = E[mx] > 0$. 然而，当未定权益空间仅仅是由所有方差有限的随机变量所组成的 Hilbert 空间的真子空间时，我们不能证明随机折现因子 m 必须是严格正随机变量。其中还需要再假定一些新的条件。不过这种数学上的细微差别对金融经济学来说，可能并不很重要。

5.7　这一节我们从另一个角度来讨论资产定价基本定理。我们考虑一个最优投资－消费问题。与以前一样，我们仍然假定只有当前和

未来两个时刻，当前是确定的，但未来则有 S 种可能；即这是所谓未定市场情形，相应的未定权益空间为 $\mathbb{R}^S$. 同时，假定市场中有 K 种基本证券，其未来价格为 S 维行向量 $x_1 \dots, x_K \in \mathbb{R}^S$; 即其支付矩阵为 $X = (x_1 \dots, x_K)^T$; 其当前价格则为 $p(x_1), \dots, p(x_K)$, 这里 p 是未定权益空间 $\mathbb{R}^S$ 上的定价函数。还要假设 $(\mathbb{R}^S, X, p(X))$ 上的无套利假设 1. 成立，即：

1. 线性定价法则成立， p 是一个 $\mathcal{M} = \mathbb{R}^S$ 上的线性函数。因此，对于任何组合 $\theta = (\theta_1, \dots, \theta_K)$, 其当前价值为

$$p(\theta \cdot X) = p(\theta_1 x_1 + \cdots + \theta_K x_K) = \theta_1 p(x_1) + \cdots + \theta_K p(x_K).$$

在这一模型下，经济活动者 (以后不妨称他为投资者) 的决策分析与确定环境下的决策分析没有本质区别。我们仍然可以通过投资者的效用函数来进行决策，只不过这一投资者对其获得的效益的判断不但要看当前价值，还要看未来的不确定的价值。也就是说，他的效用函数是当前价值和未来 S 种不同状态下的价值的函数 $u\colon \mathbb{R}_+^{S+1} \to \mathbb{R}$. 这里我们暂时也不必引进概率，而效用函数就是一个 $S+1$ 维向量空间上的函数，或者说，就是一个 $S+1$ 个变量的函数，其中第 0 个变量是投资者持有的当前价值，后 S 个变量是 S 种状态下的未来价值。 $+$ 号代表我们只考虑当前价值和未来价值非负的情形，即 $\mathbb{R}_+^{S+1}$ 是所有分量非负的 $S+1$ 维向量全体。对于这样的效用函数，我们假定下列假设成立：

强单调假设

1) $u\colon \mathbb{R}_+^{S+1} \to \mathbb{R}$ 是 $\mathbb{R}_+^{S+1}$ 上的连续函数；

2) u 对于每个变量都是严格递增的，即对于任何 $x, x' \in \mathbb{R}_+^{S+1}$, 如果 $x > x'$, 那么 $u(x) > u(x')$.

这一假设意味着，不管处于什么状态，钱总是越多越好。这里 $x > x'$ 是指 x 的每个分量都不小于 x' 的相应分量，并且 $x \neq x'$.

现在我们来考虑这一投资者的最优证券组合选择问题，即在一定的价格体系下，他应该如何来选择证券组合，使他在当前和未来的消费效用最大。这里后一句话的数学表达就是：使他的效用函数最大。这一问题与上一讲中所讨论的 Markowitz 问题当然是属于同一类型的。但是 Markowitz 问题中有收益和风险两个目标要选择。这里则只有一个目标。以后，我们将把 Markowitz 问题也纳入这样的框架。

我们还假定该投资者在当前和未来的各种状态下，都持有一定的钱 (endowment)，即假定存在非负 $S+1$ 维向量 $\omega = (\omega^0, \omega^{\mathbf{1}}) \in \mathbb{R}_+^{S+1}$, 使

得投资者选择的证券组合 $\theta \in \mathbb{R}^K$ 必须满足下列条件：

$$\begin{aligned} z^0 = \omega^0 - \theta_1 x_1^0 - \cdots - \theta_K x_K^0 &\geqslant 0 \\ z^s = \omega^s + \theta_1 x_1^s + \cdots + \theta_K x_K^s &\geqslant 0 \\ s = 1, \ldots, S. \end{aligned}$$

它们也可以简记为

$$\begin{aligned} z^0 = \omega^0 - \theta \cdot x^0 &\geqslant 0 \\ z^{\mathbf{1}} = \omega^{\mathbf{1}} + \theta \cdot x^{\mathbf{1}} &\geqslant 0 \end{aligned}$$

这里前一个条件中包含投资者在当前购买一个证券组合 θ, 因而相应的符号是负的；后一个条件中包含投资者在未来出售该组合在不同状态下所得到的价值，因而相应的符号是正的；而要求它们在各种情况下与投资者的持有之和都非负是指投资者的选择受到他所支配的资金的约束。令

$$B(\omega; x^0, x^{\mathbf{1}}) = \left\{ z \in \mathbb{R}_+^{S+1} \;\middle|\; \begin{array}{c} z^0 = \omega^0 - \theta \cdot x^0 \\ z^{\mathbf{1}} = \omega^{\mathbf{1}} + \theta \cdot x^{\mathbf{1}} \\ \theta \in \mathbb{R}^K. \end{array} \right\},$$

它称为投资者的**约束集**。投资者面临的证券组合选择问题是

$$\left\{ \begin{array}{ll} \max & u(z) \\ \text{s.t.} & z \in B(\omega; x^0, x^{\mathbf{1}}) \end{array} \right. \tag{17}$$

这一问题不一定有解。但是如果连这样的问题都没有解，那么我们就无法再进一步建立一般经济均衡理论框架。下面的定理经常被称为**资产定价第二基本定理**[2]，它主要是针对 $(\mathbb{R}^S, X, p(X))$ 上的无套利假设 2 的。

定理 5.5 在强单调假设下，问题 (17) 有解的充分必要条件为下列 $(\mathbb{R}^S, X, p(X))$ 上的无套利假设 2 成立：对于可交易未定权益 $y = \theta \cdot X$ 是正线性函数，即如果 $y = \theta \cdot X \in \mathbb{R}^S$ 满足 $y > 0$, 那么 $p(y) > 0$. 这里 $y > 0$ 表示 y 的分量非负，并且至少有一个为正。

证明 如果问题 (17) 有解 $z^* = (z^{*0}; z^{*\mathbf{1}}) \in B(\omega, x^0, x^{\mathbf{1}})$，而无套利假设 2 不满足，即存在组合 $\tilde{\theta} \in \mathbb{R}^{K+1}$ 使得

$$\begin{aligned} v^0 = -\tilde{\theta}_1 x_1^0 - \cdots - \tilde{\theta}_K x_K^0 &\geqslant 0 \\ v^s = \tilde{\theta}_1 x_1^s + \cdots + \tilde{\theta}_K x_K^s &\geqslant 0 \\ s = 1, \ldots, S, \qquad &\text{且 } (v^0, v^1, \ldots, v^S) \neq 0. \end{aligned}$$

---

[2] 有些文献中的资产定价第二基本定理是指：在完全市场中，存在唯一的无套利定价。

从而有 $z^*+v=(z^{*0}+v^0;z^{*1}+v^{*1},\ldots,z^{*S}+v^{*S})\in B(\omega;x^0,x^{\mathbf{1}})$, $z^*+v>z^*$. 因此，$u(z^*+v)>u(z^*)$. 与 z^* 是问题 (17) 的解矛盾。(以上的证明也可用普通语言表述为：如果存在套利机会，那么投资者就可不断利用套利机会增加其利益，以至不存在最优证券组合。)

反之，如果 $(\mathbb{R}^S,X,p(X))$ 上的无套利假设 2 成立，那么由资产定价基本定理 5.4, 存在 $\lambda_1,\ldots,\lambda_S>0$, 使得

$$p(x_i)=x_i^0=\lambda_1 x_i^1+\cdots+\lambda_S x_i^S,\qquad i=1,\ldots,K.$$

从而对于任何 $z\in B(\omega;x^0,x^{\mathbf{1}})$ 有

$$z^0+\lambda_1 z^1+\cdots+\lambda_S z^S=\omega^0+\lambda_1\omega^1+\cdots+\lambda_S\omega^S=\Omega.$$

这里右端的 Ω 经济含义是投资者掌握的证券组合的当前持有的价值与未来持有的折现值之和，这是个非负值(一般应该是正值)。由此可以得到，

$$0\leqslant z^0\leqslant\Omega,\quad 0\leqslant z^s\leqslant\Omega/\lambda_s,\quad s=1,\ldots,S.$$

因此，$B(\omega;x^0,x^{\mathbf{1}})$ 是个有界集。同时，它显然是一个闭集。这样，问题 (17) 变为在一个有界闭集上求连续函数的最大值问题。因此，它一定有解。 □

这条定理把投资者的最优证券组合选择问题与无套利假设联系起来。它指出，如果投资者的效用函数满足意味着“多多益善”的强单调假设，那么最优组合选择问题有解实际上与无套利假设在一定意义下是等价的。也就是说，如果无套利假设不成立，那么不可能通过一般经济均衡的框架来为证券定价。这样，无套利假设的重要性就更加突出。

下列推论还可使最优证券组合选择问题的解进一步与状态价格联系起来。

推论 设市场是完全的，u 满足强单调假设，并且可微。如果问题 (17) 有解 $z^*\in B(\omega;x^0,x^{\mathbf{1}})$, 并且 z^* 的各个分量都大于零，那么存在非负常数 ν, 使得

$$\frac{\partial u(z^*)}{\partial z^0}=\nu;\qquad \frac{\partial u(z^*)}{\partial z^s}=\nu\lambda_s,\quad s=1,\ldots,S;$$

其中 λ_s, $s=1,\ldots,S$, 是 S 个未来状态的状态价格 (即 Arrow–Debreu 证券的价格)。

证明 事实上，由上面的证明中可以看到，z^* 也是下列问题的解：

$$\begin{cases}\max & u(z)\\ \text{s.t.} & z^0-\omega^0+\lambda_1(z^1-\omega^1)+\cdots+\lambda_S(z^S-\omega^S)=0\end{cases}$$

这个问题中去除了 z 为非负向量的约束和 $\theta \in \mathbb{R}^{K+1}$ 的约束。前者是由于 z^* 的各个分量都大于零，从而约束条件 $z^s \geqslant 0,\ s = 0, 1, \ldots, S,$ 的条件不起作用。同时，由于市场是完全的，对于任何

$$z^{\mathbf{1}} - \omega^{\mathbf{1}} = (z^1 - \omega^1, \ldots, z^S - \omega^S),$$

我们总能找到适当的组合 θ^z, 使得

$$\theta^z \cdot x^{\mathbf{1}} = z^{\mathbf{1}} - \omega^{\mathbf{1}}.$$

因此，关于 θ 的约束也不起作用。这样，利用 Lagrange 乘子法，取 Lagrange 乘子为 ν, 那么 z^* 也是使 Lagrange 函数

$$u(z) - \nu[z^0 - \omega_0 + \lambda_1(z^1 - \omega^1) + \cdots + \lambda_S(z^S - \omega^S)]$$

达到最大值的解。由其各个一阶偏导数在 z^* 处为零就立即导得上述结果。 □

这一推论中增加了一些新的假设。其中之一是问题 (17) 的解要在第一卦限的内部达到。否则还要附加不等式约束，使推论中的某些等式变为不等式。另一个是完全市场假设。否则就要增加一个市场不完全的条件，即*存在* $\theta \in \mathbb{R}^K$, *使得* $\theta \cdot x^{\mathbf{1}} = z^{\mathbf{1}} - \omega^{\mathbf{1}}$. 另一方面，即使假定 u 可微，这里的 ν 还可能为零。但是如果再假定 u 是 z 的凹函数 (这也是经常作的假设，它意味着“边际效用递减”)，那么再加上强单调假设，一定有 $\nu > 0$. 这时，状态价格与“边际效用”之间就有比例关系。

上面所说的这些条件经常可以通过下列数学假设来达到：

严格拟凹性假设 *$u\colon \mathbb{R}_+^{S+1} \to \mathbb{R}$ 是严格拟凹函数，即对于任何 $x^1, x^2 \in \mathbb{R}_+^{S+1}$ 和任何 $\lambda \in (0,1)$, $u((1-\lambda)x^1 + \lambda x^2) > \min\{u(x^1), u(x^2)\}$.*

这一假设我们以后还会用到。

5.7 最后，我们举一个简单的实际应用例子来说明资产定价基本定理是如何应用的。假定有两家公司 A 和 B, 模型仍然是最简单的二期一二状态模型。这两家公司的股票价格如下表：

	当前价格	未来价格	
公司 A	62	100	30
公司 B	56	40	90

同时，在这样的条件下，有两个对两家公司的投资方案等待决策：究竟是向 A 公司投资好，还是向 B 公司投资好。

	当前投资	未来收入	
公司 A	10	10	12
公司 B	8	12	6

这个简单、而又精心构造的例子是想说明凭感觉或者用传统方法来决策是会犯错误的。很明显，一眼看去似乎应该向 A 公司投资，因为向 A 公司投资不会亏。但这都是没有考虑前面的一个表的结果。按照第一个表，**如果无套利假设成立**，那么状态价格就可确定，从而就能对投资计划的“当前价值”定价。事实上，由

$$\begin{aligned} 100p_1 + 30p_2 &= 62 \\ 40p_1 + 90p_2 &= 56 \end{aligned}$$

可解得两个状态价格分别为 $p_1 = 0.5$, $p_2 = 0.4$. 这样，对 A 公司的投资方案的**净现值** (NPV, Net Present Value) 为

$$NPV_A = 10p_1 + 12p_2 - 10 = 10 \cdot 0.5 + 12 \cdot 0.4 - 10 = -0.2,$$

$$NPV_B = 12p_1 + 6p_2 - 8 = 12 \cdot 0.5 + 6 \cdot 0.4 - 8 = 0.4.$$

因此，应该接受向 B 公司投资的方案。

当然，这个应用例子仅仅是示意性的。根据这一原理，我们不难构造出更复杂和更实用的例子来。这里，最引人注目的是：净现值的计算不是用未来价值的数学期望 (通过估计未来两个状态的实际可能发生的概率) 来计算的，而是通过由无套利假设导出的状态价格来计算的。

说明　这一讲首先指出，用“渐近无套利假设”就可导得 Ross 的 APT。这样的论证方法是 1982 年 G. Huberman 在他的学位论文 (主要结果见 [8]) 中提出的。我们在这里根据 Jarrow [10] 来介绍这种方法。类似的讨论也可在 Huang–Litzenberger [7] 的 104–108 页中找到。Duffie [4], [5] 和 Ingersoll [9] 都提到 Huberman [8]，但都没有介绍这种方法。我们在这里还指出，所谓“渐进无套利假设”就是对线性定价函数加上连续性要求。Cochrane [2] 中实际上也叙述了类似的观点。在通常的《投资学》教科书中，从来都说不清 APT 的推导。一般的做法是搞一个有限状态、有限证券的模型，然后指出，当定价不妥时就有套利机会。这其实是线性定价法则的一种“举例说明”。离 Ross 原来的 APT 还相当远。因此，本讲的第一节可以看作通常的《投资学》教科书中的“理论根据”。

实际应用的 APT 模型其实就是一个多因子定价模型。这经常会引起该取多少因子为好的问题。但是如果用随机折现因子的观点来看，任何多因子模型都等价于一个单因子 (即随机折现因子) 模型。定理 5.3 及其后面的讨论就指出了这样的结果。这里我们参考了 Cochrane [2] 的叙述。当然，纯粹从计量经济学的观点来考虑定价与有关的经济因素的关系，那是另一个问题。在资产定价的实证分析中，有时并不

拘泥其内在机理。但是从理论上来说，只要线性定价法则成立，所有的定价模型必定都是“单因子模型”。

正如我们在“代引言”中所述，无套利假设首先是在 Modigliani–Miller 1958 年的经典论文中明确提出的。在目前流行的文献中，无套利假设通常采用 Duffie [4], Ingersoll [9] 的表达形式，其中把套利机会区分为第一种套利机会 (当前价值非正，未来价值为正) 与第二种套利机会 (当前价值为负，未来价值非负)。这样一来，数学上会增加很多不必要的麻烦。我们的做法类似于 Jarrow [10] 对无套利假设的表达形式，其中明确提出“定价函数”的概念，并且通过它已经把第一种套利机会排除掉，避免了一些烦琐讨论。

资产定价基本定理的基本思想首先出现在 Ross 1976 年关于套利定价理论的经典文献 [13] 中，其中 λ_i 被冠以“状态价格”的名称。明确以两时期模型的形式的上述资产定价基本定理则出现在 Ross 1978 年发表的 [14] 中。1979 年 Harrison 和 Kreps [6] 又把这条定理赋以定理 5.4’ 那样的期望值形式。于是就可联系到概率论中的鞅的概念 (粗糙地说，鞅就是“未来”的期望值等于当前值的随机过程)。后来 Kreps [11] 又把资产定价基本定理推广到一般情形，尤其是未来价格为连续随机变量的情形。其数学证明的关键仍然在于凸集分离定理，但涉及的向量空间是无限维的。时间上的变化也逐渐从只有两个时刻发展到多个时期，直至连续变化，同样的定价思想仍然可以用更深的数学工具来演算。这种方法称为鞅方法。我们将在以后再来进一步介绍。

我们的第一个证明实际上是重复了所谓 Hahn–Banach 定理的证明。如所周知，这一定理是与凸集分离定理等价的。有关凸集分离定理的知识可参看一些数学规划理论或凸分析方面的教科书。

Duffie [5]把资产定价基本定理放到金融经济学的首要地位。请参看它的第 1 章。Ingersoll [9] 的第 2 章也是值得一读的材料。Huang–Lizenberger [7] 的有关材料在第 8 章，这是因为该书强调的是传统的“均衡定价”。令人奇怪的是 Jarrow [10] 中居然没有明确提出资产定价基本定理。

资产定价第二基本定理的说法不太普遍。有时这一名称是指“完全市场下，资产可唯一定价”。在几本常用的教科书中，只有 Duffie [5] 提到类似的结果。我们的叙述更接近于 Magill–Quinzii [12]。这一定理说明无套利假设不但是一般经济均衡的推论，甚至是一般经济均衡的前提，否则没有一个经济活动者能使其效用最大。

最后的例子取自 Copeland–Weston [3]，126 页，在那里我们还可看到其他的应用例子。

思考题和习题

1. 什么是“渐近无套利假设”？怎样理解这一假设？为什么不能用普通的无套利假设来导出它？
2. 用“渐近无套利假设”导出的 APT 是怎样的结论？它还有什么别的假设？
3. 怎样来理解和检验这里导出的 APT？

4. 什么是资产定价基本定理？它是怎样来刻划无套利假设的？怎样理解它的经济意义？

5. 怎样对“未定市场”情形利用向量空间工具来表达、理解以及证明资产定价基本定理？

6. 在未定市场情形下，什么是完全市场？它反映了一种怎样的现实金融市场的近似？它的基本性质是什么？

7. 什么叫做 Arrow–Debreu 证券？现实世界中有这样的证券吗？为什么要引进这样的证券？它们有什么经济学意义？

8. 考虑一个三期证券市场模型，其中有“今天”、“明天”、“后天”三个时刻。从今天看明天，明天可能有两个状态，而从明天的任何一个状态看后天，后天也可能有两个状态。从而，从今天看后天，后天可能有四个状态。在这样的模型中，怎样通过证券后天的价格来为它的明天定价以及为它的今天定价？这时，至少要有几个证券才能使这个证券市场为完全市场？资产定价基本定理应该怎样叙述？试根据这些不够清晰的文字描述，给出具体的数学描述，并指出，根据后天为今天无套利定价的充分必要条件是后天为明天定价以及明天为今天定价都是无套利的。

9. 怎样理解定理 5.5 的推论？这一推论也可称为：“状态价格等于边际效用之比”，它与微观经济学中的消费理论有什么相同和不同之处？它们能不能有一个共同的数学模型？

10. 你能构造一个与 5.7 节中相类似的资产定价的例子吗？用无套利定价来作投资分析是否真有道理？请阐述它的理由和不足。

参考文献

[1] Clark, S. A., 2000, Abitrage approximation theory, *Journal of Mathematical Economics*, **33**: 167–181.

[2] Cochrane, J. H., 2001, Asset Pricing, Princeton University Press, Princeton, N. J..

[3] Coperand, T. E., and J. F. Weston, 1988, Financial Theory and Corporate Policy, Addison-Wesley.

[4] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[**5**] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N.J..

[**6**] Harrison, J. M., and D. M. Kreps, 1979, Martingales and arbitrage in multiperiod securities market, *Journal of Economic Theory*, 20: 318–408.

[**7**] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[**8**] Huberman, G., 1982, A simple approach to Arbitrage Pricing Theory, *Journal of Economic Theory*, **28**: 183–191.

[**9**] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[**10**] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[**11**] Kreps, D. M., 1981, Arbitrage and equilibrium in economics with infinitely many commodities, *Journal of Mathematical Economics*, 8: 15–35.

[**12**] Magill, M. & M. Quinzii, 1997, Theory of Incomplete Markets, Vol. 1, MIT Press, Cambridge, Massachusetts.

[**13**] Ross, S. A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, **13:** 341–360.

[**14**] Ross, S. A., 1978, A simple approach to the valuation of risky streams, *Journal of Business*, 51: 453–475.

附录：数学预备知识 3

凸集和凸集分离定理：向量空间 X 中的集合 A 称为**凸集**是指

$$\forall x, y \in A, \forall \lambda \in [0,1], \qquad (1-\lambda)x + \lambda y \in A.$$

其几何意义是集合 A 中的任何两点的连接线段都在 A 中。凸集分离定理是凸集的基本性质，其几何意义是两个不相交的凸集一定可以用超平面来分离。这个性质虽然可以仅仅在向量空间的范畴中来叙述，但为了简单起见，我们假定这个 X 有 Hilbert 空间结构。X 中的超平面

H 是指 X 的“余维数为 1”的子空间的平移。余维数为 1”的子空间是指它可以表示为与 X 中的一个固定向量 μ 垂直的向量全体，即

$$L = \{x \in X \mid (\mu, x) = 0\}.$$

因此，一个超平面 H 作为 L 的平移，可以表示为

$$H = \{x \in X \mid (\mu, x - x_0) = 0\} = \{x \in X \mid (\mu, x) = \alpha\},$$

其中 x_0 是 H 中的一个固定点，$\alpha = (\mu, x_0)$. 超平面 H 把整个空间 X 分为两个半空间：

$$H_+ = \{x \in X \mid (\mu, x) \geqslant \alpha\},$$
$$H_- = \{x \in X \mid (\mu, x) \leqslant \alpha\}.$$

一个超平面把两个不相交的凸集分离是指两个凸集分别落在这两个半空间中。这就是说，如果 A 和 B 是两个不相交的凸集，那么凸集分离定理将指出，存在一个 X 中的向量 μ, 使得

$$\forall x \in A, \forall y \in B, \qquad (\mu, x) \geqslant (\mu, y),$$

其中决定超平面位置的常数 α 在这里无关紧要。上式显然等价于

$$\forall x \in A, \forall y \in B, \qquad (\mu, x - y) \geqslant 0$$

或者

$$\forall z \in (A - B), \qquad (\mu, z) \geqslant 0.$$

由于 $K = A - B$ 也是凸集，而 A 与 B 不相交等价于原点 0 不在 K 中，故 A 和 B 可用超平面分离等价于 0 与 K 的分离。

定理 1 如果 $K \subset X$ 是不包含原点 0 的闭凸集，那么存在向量 $\mu \in X$, 使得

$$\forall z \in K, \qquad (\mu, z) > 0.$$

证明 这个定理的证明与 Hilbert 空间的正交分解定理的证明思想是一样的。由于 0 不在闭凸集 K 中，故 0 到 K 之间有一正距离。与正交分解定理一样，可求出与 K 中的离 0 最近的点 μ. 于是我们有

$$\forall z \in K, \qquad \|z\| \geqslant \|\mu\|.$$

但是由于 K 是凸集，故对于任何 $z \in K$ 和 $\lambda \in (0,1)$ 也有 $(1-\lambda)\mu + \lambda z \in K$. 于是又有

$$\forall \lambda \in (0,1), \forall z \in K, \qquad \|(1-\lambda)\mu + \lambda z\|^2 \geqslant \|\mu\|^2.$$

左端展开可得

$$\|\mu\|^2 + 2\lambda(\mu, z-\mu) + \lambda^2\|z-\mu\|^2 \geqslant \|\mu\|^2.$$

因此，

$$(\mu, z-\mu) \geqslant -\frac{\lambda}{2}\|z-\mu\|^2.$$

由于上式对任何 $\lambda \in (0,1)$ 都成立。令 $\lambda \to 0$, 即得

$$\forall z \in K, \qquad (\mu, z) \geqslant \|\mu\|^2 > 0.$$

□

由此即得下列定理：

定理 2 (凸集分离定理) 如果 A 和 B 是 Hilbert 空间 X 中的两个不相交的凸集，并且 $A-B$ 是闭集，那么存在向量 $\mu \in X$, 使得

$$\forall x \in A, \forall y \in B \qquad (\mu, x) > (\mu, y).$$

六、von Neumann–Morgenstern 期望效用函数

本讲要求：了解 von Neumann–Morgenstern 期望效用函数的公理化陈述。对不确定环境下的决策问题有全面理解。了解 Allais 悖论和 Kahneman–Tversky 的研究。掌握 Arrow–Pratt 风险厌恶度量的定义及其经济涵义。对具体的期望效用函数计算 Arrow–Pratt 风险厌恶度量。初步掌握随机占优的概念。

数学预备知识：线性代数。初等概率论。微分学。

内容：

- 6.1“圣彼德堡悖论”的讨论。
- 6.2 von Neumann–Morgenstern 期望效用函数的公理化陈述。
- 6.3 Allais 悖论和 Kahneman–Tversky 的研究。
- 6.4 Arrow–Pratt 风险厌恶度量。
- 6.5 若干典型期望效用函数。
- 6.6 随机占优的概念。

在前几讲中，我们已经从无套利假设、线性定价法则和均值－方差证券组合选择出发，得到不少金融资产定价的方法，尤其是资本资产定价模型。这样定价是否合理曾经引起人们长期争论。原因在于这样的讨论被认为缺乏经济学内容，特别是几乎一直没有理会经典的一般经济均衡框架。确实，在我们前面的许多讨论中，除了个别地方我们顺便提到一些有关一般经济均衡的论证外，前几讲的主要内容是为大家介绍了一个未定权益 Hilbert 空间的理论框架。在这个框架中，经典的 Markowitz 证券组合选择理论、资本资产定价模型都被统一为线性定价法则的各种不同的表现。Ross 的 APT 方法也不过是在这一理论框架中的一个特殊问题的讨论。所有这些讨论在形式上确实也很有效，它使我们抓住了这些经典金融经济学内容的关键，即线性定价法则，或者说随机折现因子理论。然而，我们再仔细一琢磨，又会发现，这个理论框架是非常空洞的。它只是说，如果我们对未定权益全体赋以 Hilbert 空间数学结构，并且认为对未定权益的定价有线性定价法则成立，那么无论是 Markowitz 理论还是资本资产定价模型都只不过是随机折现因子的不同的表现形式。 Markowitz 理论原来就是把证券的收益率作为已知的出发点，它并没有深究这些证券为什么会有这样的收益率；而 Sharpe 的资本资产定价模型则是引进了一系列经济学假设 (投资者期望均匀性假设、无风险证券市场结清以及一般经济均衡的考虑)，从而得出资本资产定价模型中作为定价基本因素的未定权益是“市场组合”。这一番讨论虽然在开始时曾引起“β 革命”，但后来却引

起“市场组合有效性无法检验”之类的许多质疑。最后的结论仍然是回到空洞的框架：不必去管那个“市场组合”是什么，只要它均值－方差有效就行。

总之，早期的金融经济学所得到的所有结论都与金融市场中的经济活动者的行为无关。传统的观点认为商品的定价是由于经济活动者追求各自的最大利益，然后在市场上面对面地讨价还价，或是背对背地价格竞争；最后，这些经济活动者的群体行为在市场上达到均衡而形成价格体系。而套利定价论完全越过了这一过程，仅仅从市场上没有套利机会所得到的线性定价法则就得到金融商品的定价。这使得传统的经济学家们感到不好接受。下面我们开始从均衡定价论的角度来讨论有关问题，其目的是要指出，前面得到的一些结论从均衡定价的角度来看仍然成立。但是那些公式和模型并未有多少实质改变，从而，与其说给出了一种新的论证方法，不如说这是为了进一步阐明其内在的经济意义，或者说部分地解释了线性定价法则究竟是由什么来决定的。为此，我们首先要讨论经济活动者在不确定环境下的决策理论。

6.1　在带概率的不确定环境下的决策问题与概率论这门学科的历史一样古老。当时，这类问题大多是用赌博的形式提出的。既然是赌博问题，就不需要效用函数之类的概念，而只需要考虑局中人的输赢。赌博通常是可以不断重复的博弈游戏。各种事件反复随机出现使人们逐渐形成随机变量及其分布、均值 (数学期望)、方差等概念。而期望收益很自然地就成为刻划赌博输赢的总体指标。人们认为，是否值得参予一场赌博决策，在可能的情况下，可通过计算它对局中人来说的期望收益来判断。如果期望收益为正，那么对于参予者来说就是有利的，因为这虽然并不意味着你每赌必赢，但不断赌下去，其平均值将为正，即有正的收益。期望收益为负的情况恰好相反。而期望收益为零，则是意味着一场“公平赌博”。这样，期望收益的大小是“理性赌徒”用来对赌博进行决策的主要依据，或者说，期望收益可以用来给赌博“定价”。

然而，这样的决策判断很快就被质疑。这一质疑是由瑞士数学家 Nicolaus Bernoulli (1687–1759) 在整理他的伯父、概率论的奠基人之一 Jacob Bernoulli (1654–1705) 的遗著时于 1713 年向他的朋友、法国数学家 Pierre Rémond de Monmort (1678–1719) 提出的。同年，这一问题由 Monmort 在他的著作中正式发表。25 年以后，问题被 Nicolaus Bernoulli 的堂弟、当时的圣彼德堡科学院院士 Daniel Bernoulli (1700–1782) 于 1738 年解决。而这个问题后来也以“圣彼德堡悖论”而著称。

所谓圣彼得堡悖论涉及的是一场猜硬币正反面的赌博。假设第 1 次猜对，赌徒可得 2 元; 第 1 次没猜对，第 2 次猜对，赌徒可得 4 元；

前两次没猜对，第 3 次猜对，赌徒可得 8 元 $\cdots\cdots$ 一般情形是如果前 $n-1$ 次都没猜对，第 n 次猜对，赌徒可得 2^n 元。现在要问，为使一个赌徒有权参加这样的赌博，他应该先交多少钱才能使这场赌博成为“公平赌博”?

我们首先假设硬币本身是无问题的，即认为参赌者每次猜中的概率是 1/2. 这样，他第 1 次猜中的概率是 1/2，第 1 次没猜中、第 2 次才猜中的概率为 1/4，如此等等，一般为：前 $n-1$ 次没猜中, 第 n 次才猜中的概率是 $1/2^n$. 因此，参赌者可能赢的钱的数学期望应该为

$$2\cdot\frac{1}{2}+2^2\cdot\frac{1}{2^2}+2^3\cdot\frac{1}{2^3}+\cdots+2^n\cdot\frac{1}{2^n}+\cdots=\infty.$$

这就是说，参赌者无论交多少钱，这场赌博都是对他有利的。

然而，如果有人真要把这场赌博付诸实现，即使参赌者是个敢冒任何风险的地道的赌徒, 当参予这场赌博的权利的标价非常高时，他同样会踌躇不前，不愿参加。于是就产生了一个“悖论”: 为什么一场理论上的“公平赌博”，实际上却只有疯子才会愿意出任意的高价去投入?

其实这个“悖论”并非是一个 (狭义的) 科学问题，而是一个对人的行为动机的认识问题。因此， D. Bernoulli 在 1738 年作出的回答也是非科学的。他认为，人不是根据其可得的钱的数学期望来行动的，而是根据其“道德期望”来行动的。他的说法非常接近于边际效用学派的“边际效用递减”假设, 即“道德期望”并不与得利多少成正比，而与原来有多少钱有关。原来很穷，给一点钱就很满足，而原来已很有钱，要给很多钱才会增加满意程度。用数学的语言来说, “道德期望”应该是利益的导数递减的凹函数。 D. Bernoulli 所选定的“道德期望”凹函数是钱数的对数函数。这样，应该计算的不应是 x 的数学期望，而是 $a\log x$ 的数学期望，这里的 $a>0$ 适当确定。加了对数运算后，原来的变为无限大的数学期望值被代替为

$$\frac{1}{2}a\log 2+\frac{1}{2^2}a\log 2^2+\cdots+\frac{1}{2^n}a\log 2^n+\cdots=a\log 2\sum_{i=1}\frac{i}{2^i}$$

$$=2a\log 2\approx 1.39a.$$

这是一个有限值，D. Bernoulli 认为它将是人们对这场赌博愿意支付的值，其中 a 可用实际调查来确定[1]。后来还真有人作过这类调查。虽然它是因人而异的，但确实也有一定的统计稳定性.

---

[1] 上面的说法是一种简化的说法，或者说更接近于 Karl Menger 的说法。D. Bernoulli 原来的说法比较复杂。有兴趣的读者可直接参看 1954 年重新翻译发表的 D. Bernoulli 的 1738 年的论文 [3]。

D. Bernoulli 的答案虽然可再作进一步讨论，但是很可能有人会认为, 何必把这个“悖论”说得这样奥妙，这样学究气。只要把这个问题真与实际联系起来，就不会出现“悖论”。他们会说在现实中，这场赌博是不可能无限制进行下去的，总要在有限次中结束。而这个有限次数不可能很大。例如，设赌博的次数为 10 次，这时，参赌者猜对一次可能赢的钱的数学期望为

$$2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{2^2} + 2^3 \cdot \frac{1}{2^3} + \cdots + 2^{10} \cdot \frac{1}{2^{10}} = 10$$

这里假设 10 次以后再猜中，开赌者就不再付钱给参赌者。这个数字不很大，或许在不考虑道德问题时，很多人都会愿意一试；因为参赌者最多输 10 元，而如果猜得好，最多则可能赢 $2^{10} = 1024$ 元，所以这是一个风险不大的赢钱机会。再进一步说，由于社会上的总财富是一个有限值，开赌者无论如何支付不出比这个值更大的金额，这当然也限制了猜的次数。若干年以前，有人估计美国的所有财富不超过 2^{43} 美元。如果我们上面的“元”也是美元，那么最多也只能猜 43 次，于是按上面的公平赌博的定义，参赌者也只要付 43 美元，而使他至少在理论上有可能赢得全美国。因此，实际上，圣彼得堡悖论与数学中许多悖论一样，都是无限大在里面捣鬼。把这个魔鬼请出后，问题就不再有任何神秘性。

这样的回答不能说毫无道理，但它们是肤浅的回答，因为它掩盖了圣彼得堡悖论的实质。事实上，它并没有回答参赌者为什么愿意支付 10 元或 43 美元来参赌，只是因为 10 元或 43 美元对很多人来说不算是个大数目。如果世界上的总财富估计使得这场赌博有可能猜上 1000000 次，那么还有多少人愿意付上 1000000 美元来参加这场赌博呢?

D. Bernoulli 的答案则是深刻的，因为正是他首先明确指出，人们在不确定的环境下一般并不以追求直接利益的最大数学期望作为目标, 而是另有“更高的”“道德期望”。但是，D. Bernoulli 的“道德期望”长期来并不为人们所理解。一直到边际效用学派的奠基人 Carl Menger (1840–1921) 的儿子，数学家 Karl Menger (1902–1985) 把它与效用函数联系起来，尤其是 J. von Neumann (1903–1957) 和 O. Morgenstern (1902–1976) 在他们的 1944 年问世的巨著《对策论和经济行为》([24]) 中对它进行了严格的公理化论述，才开始受到数学家和经济学家的重视。

还有一点要说明的是，D. Bernoulli 的答案并不意味着有了“道德期望”以后，就消除了所有数学期望为无限大的情形。事实上，即使引进了“道德期望”，或者下面要说的“期望效用函数”，类似“圣彼德堡悖论”的情形还会出现 (参看习题 5)。他的答案只是说明，我们不必对这种数学期望出现无限大的情形感到恐惧。1959 年 Duran 曾经发表了一篇题为《成长型股票和圣彼德堡悖论》的论文 [5]，指出成长型股票的

股价估计也会出现类似“圣彼德堡悖论”那样的情形。这说明对成长型股票的估值要格外小心。

6.2 效用函数同样是在历史上 (甚至至今) 引起长期争论的概念。一个人的消费行为可以用他消费的商品量的 (效用) 函数来刻画，始终是令人怀疑的。其中效用函数无法被实际测量是它受到抨击的一个主要原因。于是后来效用函数的概念就逐渐被偏好概念所代替。但是 von Neumann–Morgenstern 在他们的书中说，如果是在带概率的不确定性的环境中，并认为人们追求的是效用函数的数学期望最大，那么在一定的合理假定下，偏好所决定的效用函数在可相差一个平移和相似 (即，相差一个仿射变换) 的意义下是唯一的。换句话说，在带概率的不确定性的环境下，效用函数实际上是可测量的，只不过这个函数的零点和单位可任意确定。就象测量温度那样，不同的零点和单位构成不同的华氏、摄氏温度计，而由此得到的不同温标可通过简单的换算互相导得。他们的书中多次提到 D. Bernoulli 的“圣彼得堡悖论”。在他们看来，D. Bernoulli 的“道德期望”其实就是效用函数的期望值。在不带不确定性的环境下，函数 x 和 $a\log x$ 所决定的偏好是一样的；从而任取一个来作为效用函数，并求其最大值，所得到的经济活动者的行为也是一样的。但是在带不确定性的环境下，如果认为经济活动者追求的是效用函数的数学期望值最大，那么这两个函数就不能都作为效用函数来看待。他们的这些主张，在当年曾被人称为是“新 Bernoulli 主义者”。典型的“新 Bernoulli 主义者”的论著可参看 Friedman–Savage [7]，其中我们可以看到，他们是怎样来反驳一些反对意见的。

在不带不确定性的一般经济均衡的讨论中，经济活动者的行为是通过对他的效用函数的最大化来决策的，这里对于消费者的效用函数是他所消费的商品量的函数，而对于生产者的效用函数是他的生产计划 (同样用商品来表示) 的函数。在带不确定性的一般经济均衡的讨论中，一种处理不确定性的办法是假定商品量都是随机变量，即它们的大小将依赖于不确定的状态。如果仍然用原来的效用函数，那么效用函数的值也将依赖于状态的随机变量，使得人们无法直接通过效用函数的值来决策。在这种情况下，人们可以希望，只要对这样的效用函数求均值 (数学期望) 以后，就能比较效用的大小。也就是说，在所涉及的随机商品量 x 集合上直接定义效用函数 u, 它应该满足下列等式：

$$E[u_1(x)] = u(x),$$

其中 u_1 是把 u 局限到非随机量时的函数，这时，$u_1(x)$ 将是随机变量。例如，具体到一个以概率 p 取 a，以概率 $(1-p)$ 取 b 的随机变量 x，这种效用函数应该满足

$$pu(a) + (1-p)u(b) = u(x).$$

其含义是一种“未定商品”的效用就等于该“未定商品”所涉及的“确定商品”的效用的均值。否则会使理论很不合理。满足这样条件的效用函数称为**期望效用函数**或 **von Neumann–Morgenstern 效用函数**。这就是 von Neumann–Morgenstern 在他们书中首先提出的。于是立即就遇到这样的问题：这样的函数是否可能合理地定义和存在。

von Neumann–Morgenstern 用数学公理化的方法处理了这一问题。他们提出的公理体系并不很复杂。尤其是我们下面要叙述的经过后人改进的体系更是简单。不过为表达这些公理，我们还需一些基本概念。假设我们考虑的经济活动者是消费者，这一消费者对他的消费活动有其自己的偏好。他的消费活动可以用他消费的商品量来刻划，但是由于现在面临的是带不确定性的环境，他面临选择的事件不再是确定的商品向量，而是它们的可能性组合。例如，“有 30% 的可能 (概率) 得到一辆自行车，有 70% 的可能 (概率) 得到一套服装。”[2]等等。消费者要对所有可能的这种组合作出其优劣判断。但是我们可从只涉及两个商品向量和一个概率的组出发来构造它们。例如，设 $L=(x,y,p)$，这里 x 和 y 是两个消费向量， $p\in[0,1]$ 是概率；整个组的含义为：消费者以概率 p 得到商品向量 x，以概率 $(1-p)$ 得到商品向量 y. 然后，再对两个这样的组 $L_1=(x_1,y_1,p_1)$ 和 $L_2=(x_2,y_2,p_2)$，构造高层次的组 $L^1=(L_1,L_2,q)$，其含义是类似的。对这种高层次的组，还能形成更高层次的组。如此等等。最终得到的实际上是一个商品向量组，其中每一个商品向量都以一定的概率被消费者所得到。所有这样的组在 von Neumann–Morgenstern 的原著中并未给出特殊的名称，但是在后来的一些文献中人们对它起了各种各样的名字；有称它为“展望 (prospect)”的 (这一名称后来被 Kahneman–Tversky [13] 拿来作为他们的理论的专门名词)，有称它为“彩票 (lottery)”的，有称它为“未定商品 (contingent commodity)”的等等。也许最后一个名称比较合适。

这样， von Neumann–Morgenstern 公理的出发点是“未定商品空间” $\mathcal{L}$，它是由“确定商品空间”中的向量 x,y 等形成形为 $L=\{(x,y,p)\}$ 的组和更高层次的组所组成。不过以后我们就以 x,y,z 等表示 $\mathcal{L}$ 中的元素，它可能是“确定商品”，也可能是“未定商品”。在这个空间 $\mathcal{L}$ 上定义了一个偏好关系 $\succsim$. 所谓偏好关系是一种排队关系。它由以下公理来刻划：

公理 0

i) (自反性) $x\succsim x$.

---

[2] 如果你感到这样的说法不好理解，那么我们也可把这种可能性组合理解为这样的“消费计划”：用 30% 的收入去买一辆自行车，用 70% 的收入去买一套服装。在数学上这不会引起任何变动。然而，在金融经济学中，我们更多地需要用概率来处理未来的不确定性。

ii) (传递性) 如果 $x \succsim y$，$y \succsim z$，那么 $x \succsim z$.

iii) (完全性) $x \succsim y$ 和 $y \succsim x$ 中至少有一个成立。

这些要求是“排队关系”的定义。数学上称为 $\succsim$ 在 $\mathcal{L}$ 上定义了一种“全序关系”。不过在这种排队关系中并非任何两个元素间都有先后，即在上述的 iii) 中，有可能两者都成立，但 $x \neq y$. 所有无先后的元素全体，在经济学上常常称它们对消费者来说“无差别”。我们可对它使用记号：$\sim$. 如果 $x \succsim y$，而 $x \nsim y$，那么我们记作 $x \succ y$.

公理 1

i) $x = (x, y, 1)$.

ii) $(x, y, p) = (y, x, 1-p)$.

iii) $((x, y, p), y, q) = (x, y, pq)$.

这里的 i) 和 ii) 是很自然的；i) 是说把确定商品向量和以概率 1 得到的商品向量看成一回事，它实际上可以看作一个定义，从而“未定商品空间” $\mathcal{L}$ 包含“确定商品空间”作为子集。但这里的 x, y 也可理解为是未定商品。ii) 是说，这里的 x, y 之间没有顺序的差别。iii) 则是说，消费者只注意它可能得到的商品向量的最终概率，而对得到它的过程不关心。这条公理中的 3 条都较自然，人们很难提出异议。

公理 2 (连续性公理) 对于任何 $x, y, z \in \mathcal{L}$,

$$\{p \in [0,1] \mid (x, y, p) \succsim z\} \quad 和 \quad \{p \in [0,1] \mid z \succsim (x, y, p)\}$$

都是闭集。

这条连续性公理一方面反映事物无大起大落变化，另一方面也是数学上的需要。否则得到的函数无连续性，处理起来很不方便。不过仔细推敲一下这条公理的含义，有可使人怀疑之处。我们讨论如下。

首先我们指出下列命题。

命题 6.1 设 $x, y, z \in \mathcal{L}$, 满足 $x \succ z \succ y$. 如果公理 0–2 成立，那么存在 $q \in (0,1)$, 使得 $(x, y, q) \sim z$.

证明 令 $A = \{p \in [0,1] \mid (x, y, p) \succsim z\}$, $B = \{p \in [0,1] \mid z \succsim (x, y, p)\}$. 那么由公理 1，$A$ 和 B 都不是空集，因为 $1 \in A$, $0 \in B$. 由公理 0，$A \cup B = [0,1]$. 由公理 2，它们都是闭集。令 $q_A = \inf_{p \in A} p$. $q_B = \sup_{p \in B} p$.

那么由 A 和 B 都是闭集，$q_A \in A, q_B \in B$. 但是又由于 $[0,1] = A \cup B$, 必须有 $q_A = q_B$, 否则 (q_B, q_A) 之间的数既不属于 A, 也不属于 B. 这是不可能的。$q = q_A = q_B$ 就是命题中所求的 $q \in (0,1)$. □

这一命题在数学上看来一切都很自然。但是如果对 x, y, z 赋以实际意义，就并非所有人都会同意。例如，设 x 是得利 101 元，y 是彻底破产，z 是得利 100 元。那么当然会认为 $x \succ z \succ y$. 但是极大多数人都不会认为肯定得利 100 元与可能冒破产的危险去争取得 101 元是“无差别”的。尽管如此，对连续性公理还是没有什么人持异议，因为上述例子是个极端的例子。既然在讨论带不确定性的问题，小概率的天灾人祸总是认为存在的，绝对可靠的事倒是不在考虑之例。因此，在理论上认为上述两件事“无差别”并无多少不妥之处。

公理 3 (独立性公理) 如果 $x \sim y$, 那么对于任何 $p \in [0,1]$ 和 $z \in \mathcal{L}$，$(x,z,p) \sim (y,z,p)$.[3]

这条独立性公理 (它也经常被称为是“替代性公理”) 是说 x 与 y 的无差别不受其它可能性的影响。表面上看来它似乎很自然，但是它受到 1988 年的诺贝尔经济学奖得主 M. Allais (1911–) 的强烈批评。关于这一点，我们放到后面介绍。

引进了公理 3 以后，我们可以有下列命题：

命题 6.2 设 $x, y \in \mathcal{L}$, $x \succ y$. 如果公理 0–3 成立，那么对于任何 $p, q \in (0,1)$, $(x,y,p) \succ (x,y,q)$ 当且仅当 $p > q$.

证明 我们先证明，对于任何 $p \in (0,1)$, $x \succ (x,y,p) \succ y$. 否则有 $(x,y,p) \succsim x \succ y$ 或者 $x \succ y \succsim (x,y,p)$. 如果 $(x,y,p) \succsim x \succ y$, 那么由命题 6.1, 存在 $\alpha \in (0,1]$, 使得 $((x,y,p),y,\alpha) \sim x$. 即 $(x,y,p\alpha) \sim x$. 令 $T = \{\beta \in [0,1] \,|\, (x,y,\beta) \sim x\}$. 那么 $p\alpha \in T$, 并且由公理 2, 它是个闭集 (两个闭集的交集)。令 $q_T = \min_{\alpha \in T} \alpha$. 则 $q_T \in (0, p\alpha]$ (因为 $x \succ y$)，并且 $(x,y,q_T) \sim x$. 但由公理 3，又有 $((x,y,q_T),y,p\alpha) \sim (x,y,p\alpha) \sim x$, 即 $(x,y,q_T p\alpha) \sim x$. 因为 $q_T p\alpha < q_T$, 这与 q_T 的定义矛盾。因此，$(x,y,p) \succsim x \succ y$ 不可能。类似地可证明 $x \succ y \succsim (x,y,p)$ 也不可能。这样，我们就证明了对于任何 $p \in (0,1)$, $x \succ (x,y,p) \succ y$.

另一方面，如果 $p > q$, 那么存在 $\alpha \in (0,1)$, 使得 $\alpha p = q$. 从而，在上面证明中的结果用 (x,y,p) 取代 x, 我们得到

$$(x,y,p) \succ ((x,y,p),y,\alpha) = (x,y,p\alpha) = (x,y,q)(\succ y).$$

反之，如果 $(x,y,p) \succ (x,y,q)$, 那么由 $(x,y,p) \succ (x,y,q) \succ y$ 和命题 6.1,

---

[3] 这条公理可放宽为只要求对 $p = 1/2$ 成立，参看 Herstein–Milnor [9] Theorem 2.

存在 $\alpha \in (0,1)$, 使得 $((x,y,p),y,\alpha) = (x,y,p\alpha) \sim (x,y,q)$. 由前面的证明，这仅当 $q = p\alpha$ 时才有可能。 □

现在我们指出，在上述假设下有以下定理:

定理 6.1 (von Neumann–Morgenstern 定理) 如果 $(\mathcal{L}, \succsim)$ 满足上述四条公理，那么在 $\mathcal{L}$ 上存在 (在至多相差一个仿射变换意义下) 唯一的期望效用函数 u，满足

$$x \succsim y \quad \text{当且仅当} \quad u(x) \geqslant u(y),$$

$$u((x,y,p)) = pu(x) + (1-p)u(y). \tag{1}$$

证明 我们先假定：在 $\mathcal{L}$ 中存在 b 和 w，使得 $\mathcal{L}$ 中的任何 x，都满足 $b \succsim x \succsim w$. 这里的 b 和 w 分别是 $\mathcal{L}$ 中的“最好的 (未定) 商品”和“最坏的 (未定) 商品”。有了这个假定以后，我们实际上只需考虑有界的效用函数。

令 $u(b) = 1$, $u(w) = 0$. 由上述假定和命题 6.1, 我们对于任何 $\mathcal{L}$ 中的 x 都可求得概率 p_x, 使得

$$x \sim (b,w,p_x) = p_x b + (1-p_x)w.$$

同时，由命题 6.2, 我们还可指出这个 p_x 是唯一的。

现在我们定义 $u(x) = p_x$. 则由公理 1 iii)，容易验证，

$$\begin{aligned}(x,y,p) &\sim ((b,w,p_x),(b,w,p_y),p) \\ &\sim (b,w,pp_x + (1-p)p_y) \\ &\sim (b,w,pu(x) + (1-p)u(y)).\end{aligned}$$

因此， (2) 成立。

还需验证这样定义的 u 是效用函数，即 $x \succsim y$ 等价于 $u(x) \geqslant u(y)$. 这是因为由命题 6.2，我们显然有 $x \succsim y$ 等价于 $p_x \geqslant p_y$.

最后，还要验证唯一性。设 u_1 是另一个满足 (2) 的效用函数。那么它必定满足

$$\begin{aligned}u_1(x) &= p_x u_1(b) + (1-p_x)u_1(w) \\ &= (u_1(b) - u_1(w))u(x) + u_1(w).\end{aligned}$$

即 u_1 与 u 只相差一个仿射变换。

现在我们来放弃存在“最好的 (未定) 商品”和“最坏的 (未定) 商品”的假定。假定我们已经对“相对最好的 (未定) 商品”b 和“相对最坏的 (未定) 商品”w 以及所有满足 $b \succsim x \succsim w$ 的 $x \in \mathcal{L}$ 证明了期望效用函数 u 的存在。但是在 $\mathcal{L}$ 中还有“更好的 (未定) 商品”$b' \succ b$ 和“更坏的 (未定) 商品”w' ($w \succ w'$) (或者只存在其中之一，对只存在其中之一的情形的证明完全一样)。于是我们又可用同样的方法得到适用于所有满足 $b' \succsim x \succsim w'$ 的 $x \in \mathcal{L}$ 的期望效用函数 u'. 这个 u' 当然与上面求得的 u 不一样，因为它满足 $u'(b') = 1$ 和 $u'(w') = 0$, 而 $u'(b) < u'(b') = 1$, $u'(w) > u'(w') = 0$. 但是这时我们可以对所有满足 $b' \succsim x \succsim w'$ 的 $x \in \mathcal{L}$, 定义

$$\tilde{u}(x) = \frac{u'(x) - u'(w)}{u'(b) - u'(w)},$$

那么它仍然是所有满足 $b' \succsim x \succsim w'$ 的 $x \in \mathcal{L}$ 的期望效用函数，并且它还满足 $\tilde{u}(b) = 1$, $\tilde{u}(w) = 0$. 由上面的证明可以立即发现，这个 $\tilde{u}$ 对所有满足 $b \succsim x \succsim w$ 的 $x \in \mathcal{L}$, 与上面得到的 u 重合。这样，我们实际上把上面得到的 u 扩充到所有满足 $b' \succsim x \succsim w'$ 的 $x \in \mathcal{L}$. 因为这样的扩充总是可能的，所以我们最后可以把 u 扩充到整个 $\mathcal{L}$ 上[4]。 □

这样，我们就在“未定商品 $\mathcal{L}$”的数学框架下，证明了 von Neumann–Morgenstern 效用函数的存在定理。这一框架对某些较复杂的经济问题的模型来说，可能还不够。但是由此出发，显然可以得到更一般的结果。例如，我们可以设法提出对于我们前面经常用到的由方差有限的随机变量所构成的“未定权益 Hilbert 空间”提出类似的结果。许多经济学家和数学家都曾对此进行深入研究。这些研究都着眼于简化定理的叙述和证明，提出一些新的公理表达形式，推广到更为一般的形式。总之，这一定理使我们对不确定性条件下的决策问题的研究有理由在一定条件下接受“期望效用函数假设”，即人们在面临带不确定性的抉择问题时，可以通过对一个效用函数求其数学期望最大来建立模型。Friedman–Savage [7] 指出，人们常从两方面来反对这一假设：一种意见是认为这不是对实际现象的一种有用或合理的解释。对此无法争辩，只能用实际的研究成果来证实。另一种意见是根本反对用一个函数来测量人们在不确定性环境下的抉择行为。对此，他们认为是不能接受的，因为这如同反对整个经济学表达理论的方式。

6.3　如上所述，von Neumann 和 Morgenstern 提出他们的效用函数公理体系的原意是企图指出，在带不确定性的环境下，人们是有可能来“测量”效用函数的。由于他们所用的数学公理化方法，这在逻辑上是滴水不漏的。如果要有什么问题，那只能到他们提出的公理体系中

---

[4] 数学上严格地说，这里用到了“超限归纳法”。

去找。而 1952 年 Allais 在一次关于 von Neumann 和 Morgenstern 的书的讨论会中提出的责难 ([1])，正是这样的问题。这迫使人们不得不进一步深思。现在我们在文献中可以看到，人们都很注意把 von Neumann–Morgenstern 效用函数的概念和通常的效用函数概念区分开来。前者常用的名称是**期望效用**。这就是说，期望效用并非是消费者真正的效用，它还包含着消费者对不确定环境的考虑，比通常的效用有更多的主观因素。这种谨慎的做法在很大程度上是 Allais 的功劳。

Allais 的贡献主要在于指出：*独立性公理不符合实际*。 Allais 构造了一个例子，并用这个例子广泛征求了意见，其中包括许多这方面的专家。结果绝大多数人都作出了与独立性公理相反的判断。连把独立性公理用另一种形式写进了他的名著：《统计学基础》(1954) ([20]，在那里作者又提出另一套描述主观概率的类似于 von Neumann–Morgenstern 期望效用函数的公理体系，由此可导出主观概率的存在性) 的经济学家和数学家 L. J. Savege (1917–1971) 都不得不承认 Allais 是有理的。从此，Allais 的这个例子就以“**Allais 悖论**”著称。

Allais 的例子十分简单。每位读者也都可以不带偏见地作出自己的判断，看是否与 Allais 所设想的一致。我们在前面讨论连续性时实际上也已提出了一个略带 Allais 悖论色彩的例子。但是在那里还能勉强自圆其说，而在这里对于独立性公理来说，则是致命的。

Allais 悖论的例子如下: 假设有以下两组事件，需要作出判断。

$$\begin{cases} A_1 & = & \text{肯定得到 100 万法郎} \\ A_2 & = & \begin{cases} \text{以 10\% 的概率得到 500 万法郎} \\ \text{以 89\% 的概率得到 100 万法郎} \\ \text{以 1\% 的概率不得利} \end{cases} \end{cases}$$

$$\begin{cases} A_3 & = & \begin{cases} \text{以 10\% 的概率得到 500 万法郎} \\ \text{以 90\% 的概率不得利} \end{cases} \\ A_4 & = & \begin{cases} \text{以 11\% 的概率得到 100 万法郎} \\ \text{以 89\% 的概率不得利} \end{cases} \end{cases}$$

对于一般人来说，在 A_1 和 A_2 之间总是会选择 A_1，而在 A_3 和 A_4 之间总是会选择 A_3. 这样的选择自然可以说出很多理由来，其中最主要的无非是要在得利和冒险之间作一番权衡。因此，在 Allais 对上百个了解概率论的人所作的调查中，绝大多数的人都作了这样的选择，连 Savege 都没有例外。

但是这样的选择恰好是违背独立性公理的。事实上，我们令

$$L_1 = (1, 1, 0.11), \quad L_2 = (5, 0, 10/11),$$

其中单位取作百万法郎。那么显然有

$$L_1 = A_1,$$

而

$$(L_2, 1, 0.11) = ((5, 0, 10/11), 1, 0.11) = A_2.$$

另一方面，我们又有

$$A_4 = (1, 0, 0.11),$$

$$(L_2, 0, 0.11) = ((5, 0, 10/11), 0, 0.11) = A_3.$$

这样，由独立性公理， A_1, A_2 之间和 A_3, A_4 之间的偏好选择问题，将完全取决于对 1 与 L_2 的选择。如果 $1 \sim L_2$，那么由独立性公理应该也有 $A_1 \sim A_2$， $A_3 \sim A_4$. 而如果认为 $A_1 \succ A_2$，则应该有 $1 \succ L_2$，从而导致 $A_4 \succ A_3$. 也就是说， $A_1 \succ A_2$ 只能与 $A_4 \succ A_3$ 相容。这就是所谓 Allais 悖论。

人们对于 Allais 悖论有各种态度。一种认为 Allais 悖论的例子太特殊，独立性公理一般来说还是成立的。于是他们还是继续用 von Neumann–Morgenstern 效用函数做各种研究。另一种态度是认为 Allais 悖论的出现是因为一般的人们都不太明事理，所以无意中作了错误判断。于是独立性公理变成了经济决策指导，而不是普通人的经济决策的反映。还有一种态度则是认为必须放弃独立性公理。那首先是 Allais 本人的态度。他认为，过去几十年来许多建立在独立性公理基础上的研究，尤其是建立在追求期望效用最大基础上的研究，都是不符合实际的，因为它们都忽略了人的心理因素对概率分布的影响。

Allais 悖论以后，又有一些学者提出各种各样的悖论，其中比较著名的有 1961 年提出的 Ellsberg 悖论 (参看例如，[6]，[15]，它其实更多地是针对 Savage 的公理系统) 等。 1970 年代起，两位心理学家 Kahneman (1934–) 和 Tversky (1937–1996) 闯进了这一研究领域。他们以心理学家特有的思路，来探讨人们在不确定环境下的决策问题，提出了许多新的观念。他们首先在很大程度上接受了 Raiffa [17] 的观点，认为对不确定环境下进行决策的分析要区别三种不同的方法。一种是**规范分析** (normative analysis)，它关注的是决策问题的理性解答，这种解答是实际决策力求迫近的。von Neumann–Morgenstern 期望效用理论就是这种分析。一种是**描述分析** (descriptive analysis)，它关注的是现实的人们实际上是怎样决策的。他们认为他们的理论更接近于这种分析，尽管在某种程度上，他们的理论也能数学公理化，但那决不是他们追求的目标。还有一种是**惯例分析** (prescriptive analysis)，它关注的是实践建议，以有助与人们可用来作出较为合理的决策。这种方法就像我们商 (管理) 学院的案例分析课那样，一般并没有系统的理论可依据，而多半

是些给人以启示的经验之谈。而“没有一种选择理论既在规范上是合适的又在描述上是精确的 (no theory of choice can be both normatively adequate and descriptively acurate)”。这样的观点与我们在第一讲中提到的“科学文化”与“文学文化”的区别有点类似。或许它可以理解为：在“科学文化”的内部还有“科学”与“文学”之分，即“规范”与“描述”之分。这就是说，我们不应该把目前对行为金融学影响极深刻的 Kahneman–Tversky 理论看作 von Neumann–Morgenstern 期望效用函数理论的敌对面，而更应该看作是一种互补面。

Kahneman–Tversky 除继承了 Allais 悖论型的例子外，还构造了许多有说服力的例子来指出 von Neumann–Morgenstern 理论的不足。这里我们列举一些这样的例子。

例 1：考虑某人携夫人去剧院看演出，他可能遇到这样的两种类似的情况：一种情况是他到了剧院时，发现已买的两张票丢了。这时他究竟是再买票 (假定还能买到几乎同样的票) 看演出，还是回家？另一种情况是他准备买票的钱丢了。这时他究竟是再掏钱买票 (假定他还有钱) 看演出，还是回家？Kahneman–Tversky 发现，大多数人在第一种情况下会选择回家；而在第二种情况下会选择再买票。对这个例子的结果我们可以说出许多心理上的原因，但是你很难用期望效用函数来刻划，因为用钱来衡量时，丢票与丢钱的结果是一样的。

例 2 (见 [22])：假定有某个遭受疫情威胁的村子。这个村子有 600 人。又假定有两个援救方案：第一个方案是使 400 人死，200 人获救；另一个方案是 600 人全死的概率是 2/3, 600 人全获救的概率是 1/3. 这两个方案的获救人数的期望值是一样的。按期望效用函数的理论，由于一般人都厌恶风险，通常都会选择第一种方案。但是由于这个问题涉及人们的生死，人们的实际选择很大程度上取决于怎样来叙述这两个方案。在一次调查中，这两个方案被叙述为：第一方案是 200 人获救；第二方案是 600 人获救的概率是 1/3, 没有人获救的概率是 2/3. 结果赞成第一方案的人数与赞成第二方案的人数的比例为 72:28 (被调查总人数 152，实际人数的比例为 109:43)。在另一次调查中，又把援救方案叙述为：第一方案是 400 人死去，第二方案是使全体获救的概率是 1/3, 600 人死去的概率是 2/3. 结果赞成第一方案的人数与赞成第二方案的人数的比例竟反过来变成 22:78 (被调查总人数 155，实际人数的比例为 34:121)。

这两个例子很能说明 Kahneman–Tversky 的思想。如果说，Allais 悖论中只涉及抉择人的“利益”和不确定性，使得抉择人作为“社会人”的特征还不很明显，那么 Kahneman–Tversky 的例子则完全突出了抉择人的“社会人”特征。用 Kahneman–Tversky 的术语来说，他们认为人

们在不确定环境下的抉择，很大程度上取决于“决策的框架 (framing of decision)”。第一个例子中虽然都只涉及两张剧院票的价值，但是两张票本身与两张票的价值对抉择人来说，仍然是不一样的因素。第二个例子中完全一样的方案，就因为从获救还是从死亡的不同感受来考虑，人们的抉择也会不同。这里涉及的钱、剧、生、死等都与“社会人”特征有关。所谓不同的“决策的框架”只不过是在框架中要强调哪些特征。而从我们在第一讲中所提到的逻辑角度来考虑，那么或许我们可以说，Kahneman–Tversky 的例子实际上指出，在涉及备选方案的内涵时，人们经常使用的是内涵逻辑，而不是外延逻辑。外延上一样的对象，由于人们理解的内涵不同，就会有不同的逻辑推理结果。所谓不同的“决策的框架”，实际上是在对同一个 (外延) 对象赋以不同的内涵。从这个角度来看，我们也能理解 von Neumann–Morgenstern 期望效用函数理论的局限。这种理论是从干巴巴的公理体系出发的，这些公理已经不食人间烟火，而变为纯粹的数学符号。但是活生生的人在作抉择时决不会用纯粹的数字与符号来思考。他们会想到钱、剧、生、死等等对自身的许多无法言表的含义，而这些含义是不可能完全体现在那些公理中去的。

或许也就因为这样，Kahneman–Tversky 的“展望理论 (prospect theory)”更多地被强调为是一种描述分析，使人们在应用这种理论时要多考虑活生生的现实世界，而不要只沉浸在干巴巴的数学符号中。因此，在他们的分析中，他们经常运用许多实验案例来说明问题，其中尤其是用实验案例来提出一系列与期望效用函数理论相悖的“效应”。这些“效应”有**确定性效应** (certainty effect) (与肯定结果相比低估可能发生的结果)、**反射效应** (reflection effect) (对增益与损失的反应犹如镜像，即低估增益，高估损失)、**孤立效应** (isolation effect) (对同一事件的不同片面强调，会导致绝然相反的偏好) 等等。它们都是人们在运用期望效用函数理论时需要注意的。

展望理论最后体现为期望效用函数的一种修正。仍然运用上面讨论的期望效用函数的理论框架，关于未定商品 (x, y, p), 对期望效用函数 u 来说，

$$u((x, y, p)) = pu(x) + (1 - p)u(y),$$

但是对“Kahneman–Tversky **值函数** (value function)” v 来说，

$$v((x, y, p)) = \pi(p)v(x) + \pi(1 - p)v(y),$$

其中作为单变量函数的值函数 v, 不再像效用函数一般是凹函数，而是对自变量的正值 (“增益 (gain)”) 是凹函数，对自变量的负值 (“损失 (loss)”) 则是凸函数； π 是**权重函数** (weighting function), 它是对概率的一种估计，通常把低概率高估，把高概率低估。两者的图象如图 1。

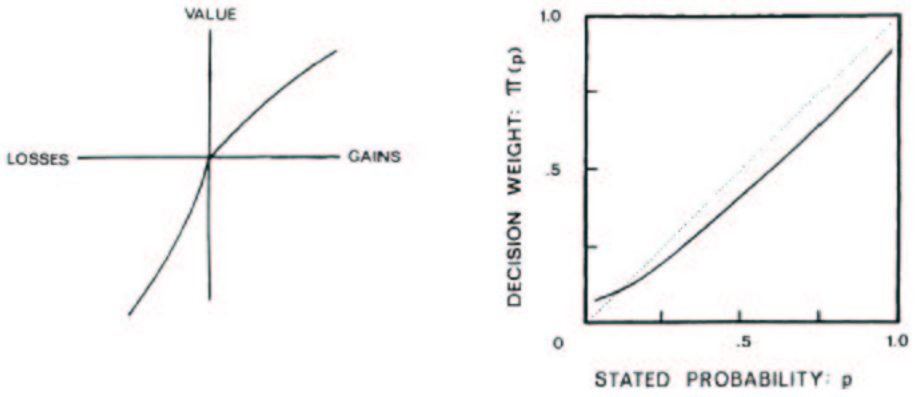


图1: Kahneman–Tversky 值函数 (左) 与权重函数 (右)。引自 [13]

Kahneman–Tversky 的研究引起了经济学界的极大重视，尤其是金融经济学研究的极大重视。金融经济学的一个新研究方向：行为金融学因此而应运而生。许多这方面的学者用他们的理论来解释金融市场的反常现象 (即不符合经典金融经济学的现象，例如股市的过度反应) 等，在一定程度上获得成功。Kahneman 也因此与实验经济学的前驱者 V. L. Smith (1927–) 一起荣获 2002 年的诺贝尔经济学奖。而早逝的 Tversky 则成了继 Black 以后的又一位在诺贝尔经济学奖公告上被明确表彰的已故学者。

尽管如此，直到现在为止，在带不确定性的经济学问题的研究中，承认 von Neumann–Morgenstern 效用函数的还是主流。Allais 的后继者虽然在放弃独立性公理等方面作了很多很深入的研究，但是由于其理论上过于复杂，始终没有达到取代 von Neumann–Morgenstern 效用函数的地步。我们在 Fishburn [6] 这样的专著中可以找到许多从各方面来推广期望效用函数理论的公理体系的研究 (各种非线性、非可加、非传递效用)。但形式推广只是一种干巴巴的数学，而问题则在于活生生的人。因此，建立在大量心理实验基础上的 Kahneman–Tversky 理论虽然在数学上并不高深，甚至仅仅是期望效用函数的一种简单“修正”，反而成了人们关注的焦点。它在某种意义下，变成了对期望效用函数理论的一种新的肯定。这样，我们仍然应该把期望效用函数理论作为金融经济学的一份宝贵财富，谨慎地把它保存下来。

6.4 现在我们继续讨论期望效用函数理论。设某经济活动者的期望效用函数为单变量函数 $u(x)$. 我们不妨设这里自变量的含义就是收入。假设 $x, y \geqslant 0$ 为两种可能的收入；得到 x 的概率为 p，而得到 y 的概率为 $(1-p)$. 那么由 von Neumann–Morgenstern 效用函数的定义，并

仍用原来的记号，可得这一事件的效用为

$$u((x,y,p)) = pu(x) + (1-p)u(y).$$

此人对 (x,y,p) 这一事件中所包含的风险的态度可由这个值与 $u(px+(1-p)y)$ 的比较来刻划。如果

$$u(px+(1-p)y) = u((x,y,p)),$$

那么称该经济活动者为**风险中性者**。如果

$$u(px+(1-p)y) > u((x,y,p)),$$

那么称该经济活动者为**风险厌恶者**。如果

$$u(px+(1-p)y) < u((x,y,p)),$$

那么称该经济活动者为**风险爱好者**。要注意的是，这些提法都依赖于 x, y 和 p.

上述这些提法可以这样来理解：一般总认为 u 是收入的连续递增函数。于是对于风险厌恶者来说，使 $u(x') = u((x,y,p))$ 的 x' 应该比 $px+(1-p)y$ 小 (如图 1)，而对于风险爱好者来说，应该比 $px+(1-p)y$ 大；而对于风险中性者来说，两者则总相等。

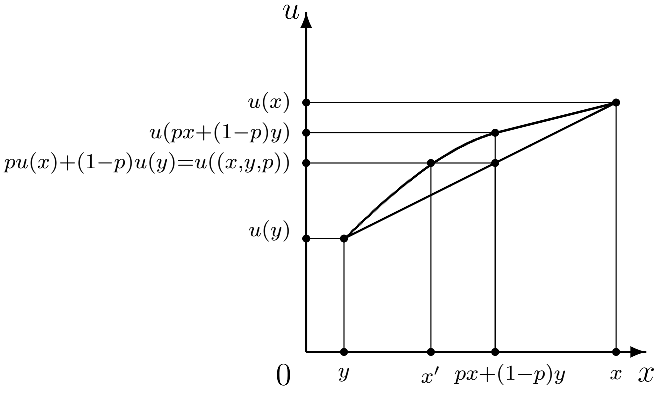


图 2: 单变量效用函数的图象

这样说，还是比较抽象。我们再进一步把 x 解释为正常收入，即可得到它的概率 p 较接近于 1. 如果 y 是一场灾难，即 y 远小于 x，那

么 $x - x'$ 可以看作经济活动者为灾难 y 愿意支付的保险费用。这时我们可以看出：风险厌恶者愿意支付较高的费用，而风险爱好者则不愿意。反之，如果 y 是一笔巨大的彩票收入，y 远大于 x，那么 $x' - x$ 可以看作经济活动者愿意购买的彩票价格。这时有：风险爱好者愿意购买价格较高的彩票，而风险厌恶者则不愿意。这两个例子足以说明上面这些名称的含义.

那么怎样来刻划厌恶风险的程度呢? 如果从函数的图象来看 (再参看图 2)，自然是曲线向上弯得越厉害，对风险就越厌恶。曲线的弯曲程度可以用函数的二阶导数来刻划，但光是二阶导数还不行，因为它不是对仿射变换不变的量。虽然二阶导数不受效用函数换零点的影响，但是要受效用函数换单位的影响。从而不同的人之间的风险厌恶程度就无法比较。为了使这个量不受换单位的影响，Arrow [2] 和 Pratt [16] 建议风险厌恶程度用

$$A_u(x) = -\frac{u''(x)}{u'(x)}$$

来衡量。这就是所谓 **Arrow–Pratt 风险厌恶度量**，或者 Arrow–Pratt **绝对风险厌恶函数**。 $T_u(x) = 1/A_u(x)$ 则称为 **风险容忍函数**，而 $R_u(x) = xA_u(x)$ 则称为**相对风险厌恶函数**。

为了理解这一度量的意义，我们需要引进**函数平均**的概念。设 $a_1, a_2, \ldots, a_n > 0$ 为 n 个正数。通常的算术平均定义为

$$\frac{a_1 + a_2 + \cdots + a_n}{n}.$$

平方平均定义为

$$\left(\frac{a_1^2 + a_2^2 + \cdots + a_n^n}{n}\right)^{1/2}.$$

几何平均定义为

$$\sqrt[n]{a_1 a_2 \cdots a_n} = \exp\left(\frac{\log a_1 + \log a_2 + \cdots + \log a_n}{n}\right).$$

对于一般的定义域为 $(0, +\infty)$ 的严格单调函数 f 定义的 f-**平均**为：

$$M_f(a) = f^{-1}\left(\frac{f(a_1) + f(a_2) + \cdots + f(a_n)}{n}\right),$$

这里 f^{-1} 表示 f 的反函数。

回忆单变量凸函数的定义，我们可以看到，对于凸函数来说，有

$$f\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right) \leqslant \frac{f(a_1) + f(a_2) + \cdots + f(a_n)}{n}.$$

如果 f 是严格单调函数，那么对上式两端作用 f^{-1}，它也可表达为：(严格单调) 凸函数 f 就是 f-平均不小于算术平均的函数。对于凹函数 u 则可以说，(严格单调) 凹函数 u 就是 u-平均不大于算术平均的函数。

如果对于任何正数向量 $a = (a_1, a_2, \ldots, a_n)$, a 的 g-平均总不小于 f-平均，那么当 f 的值域是 $(0, +\infty)$ 时，不难验证 g 与 f^{-1} 的复合函数 $h = g \circ f^{-1}$ 的 h-平均总不小于算术平均。因此， $h = g \circ f^{-1}$ 是凸函数。如果它二阶可导，那么 $h'' \geqslant 0$. 由复合函数和反函数求导法则，可以求得

$$h'(x) = (g(f^{-1}(x))' = g'(f^{-1}(x))(f^{-1}(x))' = \frac{g'(f^{-1}(x))}{f'(f^{-1}(x))}.$$

$$h''(x) =$$

$$\frac{g''(f^{-1}(x))(f'(f^{-1}(x))^{-1}f'(f^{-1}(x)) - g'(f^{-1}(x))f''(f^{-1}(x))(f'(f^{-1}(x))^{-1}}{(f'(f^{-1}(x))^2}$$

$$\geqslant 0.$$

因此，

$$\frac{f''(f^{-1}(x))}{f'(f^{-1}(x))} \leqslant \frac{g''(f^{-1}(x))}{g'(f^{-1}(x))}.$$

由 x 的任意性，即得

$$\frac{f''}{f'} \leqslant \frac{g''}{g'}.$$

这样 f''/f' 可以看作 f 的凸的程度的度量，$-u''/u'$ 就可看作 u 的凹的程度的度量。这就是 Arrow–Pratt 绝对风险厌恶度量的数学意义。

把上述结果写成数学形式，即我们有下列命题：

命题 6.3　设 f, g 为定义在 $(0, \infty)$ 上和取值在 $(0, \infty)$ 上的光滑严格单调函数。那么,

$$\forall x, y \in (0, \infty), \quad \forall p \in [0, 1],$$

$$f^{-1}(pf(x) + (1-p)f(y)) \leqslant g^{-1}(pg(x) + (1-p)g(y)) \tag{2}$$

等价于

$$\forall x \in (a, b), \quad \frac{f''(x)}{f'(x)} \leqslant \frac{g''(x)}{g'(x)} \tag{3}$$

联系关于资产定价基本定理的讨论，我们可以看到，如果把资产的价格看作一个期望效用函数，那么定理 5.4' 实际上是说，存在一个

等价概率测度，使得这个期望效用函数是风险中性的。因此，文献中也经常把套利定价称为**风险中性定价**。

6.5　以下我们来介绍一些常用的期望效用函数。在上一讲讨论资产定价第二基本定理 5.5 时所涉及的效用函数与概率无关，因而也没有涉及期望效用函数。但是如果 S 个未来状态的概率分别为 $q_1, \ldots, q_S > 0$, $(q_1 + \cdots + q_S = 1)$, 那么一种常用的效用函数为

$$u(z^0, z^1, \ldots, z^S) = v(z^0) + \delta \sum_{s=1}^{S} q_s v(z^s),$$

其中 v 为定义在 $\mathbb{R}_+$ 上的严格递增连续凹函数，$\delta \in (0,1)$ 代表“折现因子”。这时，u 就可看作期望效用函数。

形式为

$$u(x) = \frac{1-\gamma}{\gamma}\left(\frac{ax}{1-\gamma} + b\right)^{\gamma}, \qquad b > 0, \gamma > 0,$$

的效用函数称为**双曲绝对风险厌恶** (hyperbolic absolute risk aversion, HARA) 类效用函数。其特点在于其风险厌恶度量 $r_u(x) = -u''(x)/u'(x)$ 满足

$$T_u(x) = \frac{1}{A_u(x)} = \frac{x}{1-\gamma} + \frac{b}{a}.$$

这是一个线性函数，因而 HARA 类函数也经常被称为**线性风险容忍** (linear risk tolerance, LRT) 类函数。这类函数的特点为其风险厌恶程度的绝对值随着 x 的绝对值增大而减少 (“钱越多就越对风险不在乎”)。但是其符号还取决于 γ 是 < 1 还是 > 1; 前者对应“钱越多就越不怕风险”；后者对应“钱越多就越不想冒险”。由于这里为保证幂函数有定义，还有一个约束条件：$b + ax/(1-\gamma) > 0$, 这使得 x 对 $\gamma < 1$ 有下界，对 $\gamma > 1$ 有上界。因此，如果允许 x 取负值，后一句话对极限情况来说应该是“赔得越多就越不在乎风险”。

$\gamma = 1$ 对应 $u(x) = ax$. 这是风险中性情形。$\gamma = 2$ 是二次函数情形。当 $b = 1$, 令 $\gamma \to -\infty$, $u(x) \to -e^{-ax}$. 令 $b = 0$, $\gamma < 1$, u 是幂函数，这时相对风险厌恶函数 R_u 是常数 $1 - \gamma$. $u(x) = \log x$ 也可以看作 HARA 函数的一种，这是因为 $\log x = \lim_{\gamma \to 0}(x^\gamma - 1)/\gamma$. 因此，它在某种意义下，可看作 $\gamma = 0$ 的 HARA 函数。其相对风险厌恶函数 $R_u = 1$. 由此可见，HARA 效用函数类是相当广的函数类，它适用于许多金融问题的讨论。

6.6　用期望效用函数来判断两个随机事件 (未定商品) 之间的优劣是仅仅针对单个经济活动者而言的。于是很自然地可提出这样的问

题，能否对经济活动者的群体来提出同样的问题？如果仍然用期望效用函数来讨论，这意味着对一类期望效用函数来提出同样的问题。在数学上，它可以形成这样的一个问题：设 X 是一个随机变量的集合。U 是一个单变量函数集合。在 X 中定义了一种半序关系 $\succcurlyeq$: 对于任何 $x, y \in X$,

$$x \succcurlyeq y \iff \forall u \in U, \quad E[u(x)] \geqslant E[u(y)].$$

这里 x,y 的含义可以是两种风险证券的收益率，而 u 的含义可以是某个投资者的期望效用函数，而 U 的含义可以是某个投资者群体的期望效用函数全体。从而，这样定义的半序关系意味着这群投资者都认为后一种证券不比前一种证券差。这样的概念对金融经济学的意义是显而易见的。这里需要注意的是：一般来说，这种关系是一种半序关系，而不是全序关系；也就是说，并不是任何两种风险证券都可比较优劣的。同时，这不过是用期望效用函数理论来刻划投资者群的偏好，这样的讨论必定受到期望效用函数理论的局限。在 Markowitz 理论中，“收益大，风险小”也是对证券的收益率定义了一种半序关系。遗憾的是，这种半序关系并不能简单地归结为上述情形。对此，我们以后再进行讨论。

对于金融学问题来说，上面的随机变量集合 X 似乎应该取未定权益 Hilbert 空间 $\mathcal{M}$. 但是由于目前的问题不涉及定价之类的内容，人们经常讨论的 X 就是 (实数轴上) 的随机变量全体。上述的半序关系就被称为**随机占优** (stochastic dominance)。效用函数集合 U 则可取这样两种：1. 单调不减函数类；2. 单调不减凹函数类。而相应的随机占优也被称为**一阶和二阶随机占优**。如果把随机变量理解为证券 (或更一般的未定权益) 的收益率，那么这两类函数可分别理解“收益越多越好”和“风险厌恶”。有时还需要讨论**三阶随机占优**，它是指 U 为有非负三阶导数的递增凹函数类。绝对风险厌恶递减的效用函数属于这一类，但它还包含更多的函数。

关于随机占优的主要结果是以下两条定理。但是为了避免一些数学上的困难，我们假定 X 为所有只取有限区间中的值的随机变量这样，它的数学期望、方差以及对一般的效用函数的期望效用总存在。这一假定对于实际应用足够，并且也不难用一些常规的数学技巧把这一假定取消。同时，我们还要注意到

1. 对于一阶随机占优的效用函数类为非负可测函数的积分，即 $u(t) = u(0) + \int_0^t u'(s)ds$, 且 $u' \geqslant 0$.
2. 对于二阶随机占优的效用函数类为上述函数类中的一个子类，其中的函数 u 的导数 u' 是非正可测函数的积分，即 $u'(t) = u'(0) + \int_0^t u''(s)ds$, 且 $u'' \leqslant 0$.

定理 6.2 (一阶随机占优定理) 设 x, y 为两个只取有限区间中的值的随机变量， $F(s) = \mathrm{P}\{x \leqslant s\}$ 和 $G(s) = \mathrm{P}\{y \leqslant s\}$ 分别为它们的累积分布函数。那么 x 一阶随机占优于 y 的充分必要条件为

$$\forall s \in \mathbb{R}, \qquad F(s) \leqslant G(s). \tag{4}$$

证明 所谓 x 一阶随机占优于 y 是指对于上述函数类中的任何 u 有

$$E[u(x)] \geqslant E[u(y)], \quad \text{即} \quad \int_{-\infty}^{\infty} u(s)dF(s) \geqslant \int_{-\infty}^{\infty} u(s)dG(s).$$

但由分部积分法

$$\int_{-\infty}^{\infty} u(t)d(F(t) - G(t)) = u(t)(F(t) - G(t))\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (F(s) - G(s))du(s)$$

$$= -\int_{-\infty}^{\infty} (F(s) - G(s))du(s) = -\int_{-\infty}^{\infty} (F(s) - G(s))u'(s)ds.$$

其中我们要注意到，由于 $F - G$ 实际上只在一个有限区间中不为零，上述的积分其实都是只在有限区间中进行的。这一等式对于任何非负可测函数 u' 成立。考虑到随机变量的累积分布函数都是右连续递增函数，容易证明，最后一个表达式非负的充要条件为 (4)。 □

由于 $u(s) = s$ 是满足上述要求的单调不减函数，故 x 一阶随机占优于 y 的一个必要条件为：$E[x] \geqslant E[y]$. 从而，对于未定权益的收益率的比较，一阶随机占优首先意味着期望收益率上的“占优”。另一方面，我们还要注意到，累积分布函数的特点都是从 $-\infty$ 处的 0 变到 ∞ 处的 1. 因此，一个累积分布函数处处都不大于另一个累积分布函数，意味着它们的图象是夹在纵坐标的 0 与 1 之间的带形中的从 $-\infty$ 处的 0 单调变到 ∞ 处的 1 的两条曲线，其中一条始终不越过另一条。

定理 6.3 (二阶随机占优定理) 设 x, y 为两个只取有限区间中的值的随机变量， $F(s) = \mathrm{P}\{x \leqslant s\}$ 和 $G(s) = \mathrm{P}\{y \leqslant s\}$ 分别为它们的累积分布函数。那么 x 二阶随机占优于 y 的充分必要条件为

$$\forall t \in \mathbb{R}, \qquad \int_{-\infty}^{t} F(s)ds \leqslant \int_{-\infty}^{t} G(s)ds. \tag{5}$$

证明 令

$$S(s) = \int_{-\infty}^{s} (F(s) - G(s))ds.$$

则 S 只在有限区间中不为零，并且 $dS(s) = (F(s) - G(s))ds$. 由上一定理的证明，再运用分部积分法，可得

$$\int_{-\infty}^{\infty} u(t)d(F(t) - G(t)) = -\int_{-\infty}^{\infty} (F(s) - G(s))u'(s)ds$$

$$= -S(s)u'(s)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} S(s)u''(s)ds = \int_{-\infty}^{\infty} S(s)u''(s)ds.$$

由于这里的 u'' 可取任何非正可测函数，因而上述等式的右端非负的充要条件为 (5)。 □

对于二阶随机占优来说，同样由于 $u(s) = s$ 是满足要求的单调不增凹函数， $E[x] \geqslant E[y]$ 仍然是 x 二阶随机占优于 y 的必要条件。但是由 x 二阶随机占优于 y, 却不能导出 x 的方差比 y 的方差小。也就是说二阶随机占优并非是 Markowitz 均值－方差准则的充分条件。事实上，我们很容易构造反例来说明这一点。下列反例取自 Hanoch–Levy [9]：设随机变量 x 和 y 只取 $0, 1, 2$ 三个值，但所取的概率不同。x 取这三个值的概率为：0.65, 0.25, 0.10; 而 y 取这三个值的概率为：0.86, 0.89, 0.11. 于是我们有

$$\begin{array}{llll}
F(0) = 0.65 & G(0) = 0.86 & F(0) - G(0) = -0.21 & \int_{-\infty}^{0}(F - G)ds = 0 \\
F(1) = 0.90 & G(1) = 0.89 & F(1) - G(1) = 0.01 & \int_{-\infty}^{1}(F - G)ds = -0.21 \\
F(2) = 1.00 & G(2) = 1.00 & F(2) - G(2) = 0 & \int_{-\infty}^{2}(F - G)ds = -0.20
\end{array}$$

因为对任何 t, 有 $S(t) = \int_{-\infty}^{t}(F - G)ds \leqslant 0$, 故由定理 6.3, x 二阶随机占优于 y. 但是 $E[x] = 0.45 > E[y] = 0.25$, $\mathrm{Var}[x] = 0.4475 > \mathrm{Var}[y] = 0.4075$. [5]

另一方面，我们也可构造反例说明，二阶随机占优也不是 Markowitz 均值－方差准则的必要条件。另一个反例如下：设随机变量 x 以 0.80 的概率取 1, 以 0.20 的概率取 100; y 以 0.99 的概率取 10, 以 0.01 的概率取 1000. 则 $E[x] = 20.8 > E[y] = 19.9$, $\mathrm{Var}[x] = 1568.16 < \mathrm{Var}[y] = 9702.99$. 即 x 均值－方差优于 y, 但对于递增凹函数 $u(s) = \log_{10} s$ 来说，$E[u(x)] = 0.4 < E[u(y)] = 1.02$. 即 x 并未二阶随机占优于 y.

那么 Markowitz 均值－方差准则能否也可通过某一类随机变量或某一类效用函数类来定义呢？对此，Tobin [21] 曾有过两个建议。1. 运用二次效用函数；2. 假定随机变量服从正态分布。对于第 1 点，如果还要求效用函数是单调不减凹函数，那么这个函数的二次项系数必须

---

[5] 同时，我们还可看到 F 并不总不大于 G, 故由定理 6.2, x 并不一阶随机占优于 y. 因此，这个例子也说明二阶随机占优或数学期望较大并不是一阶随机占优的充分条件。

是负的，并且当它达到最大值后，最大值点右边的“下降抛物线”就不能采用。如果不考虑这部分“下降抛物线”，那么可以证明，均值－方差准则是对这类效用函数随机占优的充分条件；即如果 $E[x] \geqslant E[y]$, $\mathrm{Var}[x] \leqslant \mathrm{Var}[y]$, 那么对于任何 $u(t) = -at^2 + bt + c$, $a \geqslant 0$, 在 u 为递增凹函数的区间内，有 $E[u(x)] \geqslant E[u(y)]$ (习题)。然而，反之仍然不成立，即均值－方差准则并非是对这类效用函数随机占优的必要条件，因为前面的第一个反例在这种情形仍然有效 (我们可以考虑在 $[0, 2]$ 上的递增凹二次效用函数)。对于第 2 点，情况仍然类似。这时，我们考虑有 Taylor 展开的效用函数。由于标准正态分布随机变量的奇阶矩都为零，偶阶矩都是二阶矩的幂与正常数的乘积，任何单调不减凹函数的期望效用都类同于一个二次效用函数的期望效用。因此，均值－方差准则对正态分布的随机变量来说是二阶随机最优的充分条件 (某些数学细节尚需证明)。但反之还是不成立。这只需考虑一个 y 的累积分布函数处处大于 x 的累积分布函数的例子即可。由定理 6.2 和 6.3, 这时 x 一阶、二阶随机占优于 y, 但并不能导得 x 均值－方差优于 y, 相反，我们可以使 x 的方差任意大。总之，尽管 Markowitz 本人曾经尽力想使他的均值－方差准则纳入期望效用函数的框架，如上所述，均值－方差准则实际上是不可能完全纳入这一轨道的。正如我们前几讲中所已经阐明的， Markowitz 理论其实只是线性定价法则的一种表达形式，而线性定价法则并不需要期望效用函数理论。

说明 关于“圣彼德堡悖论”和 von Neumann–Morgenstern 效用函数基本上重复了作者以前的论述 [25]。这些内容都是相当经典的，几乎每一本微观经济学的教科书都要讲述它们。我们也可直接参考经典文献 Bernoulli [2] 和 von Neumann–Morgenstern [24]。比较深入、全面的简述可在 Kreps [15] 中找到。这方面的专著也不少。Fishburn [6] 就是一本总结相当全面、观点发人深思的好书，其中也包含不少有关 Kahneman–Tversky 理论的内容。至于 Kahneman–Tversky 的所谓“展望理论 (Prospect Theory)”，可直接参看他们的经典论文 [13]，这篇论文据说已经成为 Econometrica (计量经济学杂志) 有史以来被引证次数最多的论文。尤其是它已成为行为金融学的基础。对于投资学来说，他们的理论的具体体现可参看 Kahneman–Riepe 于 1998 年发表的一篇很有意思的论文《投资者心理学的几个方面》([14])。

关于 von Neumann–Morgenstern 期望效用函数理论，我们的叙述形式基本上引自 Varian [21]，但已根据原始文献 Herstein–Milnor [9] 进行了补充。其他是根据一些史料而感发的我们自己的理解。Huang–Linzenberger [10], Ingersoll [11], Jarrow [12] 也都有有关论述，其中以 [10] 的叙述最详尽，[12] 最简略。但是通常都只讨论很特殊的离散情形。一般的概率空间上的 von Neumann–Morgenstern 效用函数的存在性问题相当复杂。[6] 的第一章的最后有这类存在定理的两个版本。

Arrow–Pratt 风险厌恶度量作为描述效用函数的弯曲程度的“仿射不变量”已经被广泛应用。命题 6.3 指出了它与函数的凸性之间的关系。这一引理出现在 Pratt [16] 中，那里还有一些刻划函数凸性的其他条件。Ingersoll [11] 的第 1 章中的定理 5

也是类似的结果。在许多文献中出现的 HARA 效用函数类也可在 Ingersoll [11] 的第 1 章中找到。

随机占优问题在 Huang–Linzenberger [10] 中用整整一章来进行讨论。Ingersoll [11] 则仅仅简单提及。Jarrow [12] 则完全没有提到。我们的叙述主要根据 Hanoch–Levy [8]，也比较接近于 Ingersoll [11] 中的叙述。需要注意的是，我们的二阶随机占优的定义与 [10] 中不同。在 [10] 中，二阶随机占优首先要求两个随机变量的数学期望相同。这样的定义是 Rothschild–Stiglitz [18], [19] 所提出的。它可使得二阶随机占优是均值－方差准则的充分条件。在那里，作者也证明了一个随机变量可表示为一个二阶随机占优于它的随机变量与一个“噪声”的和。 Rothschild–Stiglitz 的这两篇重要工作，对信息经济学的发展有很大影响。

思考题和习题

1. 什么是“圣彼德堡悖论”？你对它有什么看法？
2. 什么是 von Neumann–Morgenstern 效用函数？它的主要性质是什么？它是根据哪几条公理建立起来的？
3. 有了 von Neumann–Morgenstern 效用函数以后，“圣彼德堡悖论”是否就不会发生？试考虑一个当 $x \to +\infty$ 时，$u(x) \to +\infty$ 的期望效用函数，并由此构造一场用 u 来衡量收益的赌博，它仍然有“圣彼德堡悖论”的特征。
4. 你对 Allais 悖论的例子怎样判断？为什么说它与“独立性公理”矛盾？
5. “肯定得 800 元”与“以 85% 的概率获得 1000 元”两者你选择哪一个？有人选择前者，这说明他的期望效用函数有什么特点？
6. 求下列期望效用函数的 Arrow–Pratt 绝对风险厌恶函数：

   $$u(x) = x - \frac{b}{2}x^2, \quad -e^{-bx}, \quad b\log x,$$

   其中 $b > 0$ 是常数。
7. 设 u 是满足 $u'(x) > 0$, $u''(x) < 0$ 的期望效用函数。证明，如果它的绝对风险厌恶函数 A_u 满足 $A'_u(x) \leqslant 0$, 那么 $u'''(x) > 0$.
8. 什么是随机占优？它有什么经济学含义？
9. 设随机变量 x 和 y 满足 $E[x] \geqslant E[y]$, $\mathrm{Var}[x] \leqslant \mathrm{Var}[y]$, 试证，对于任何 $u(t) = -at^2 + bt + c$, $a \geqslant 0$, 在 u 为递增凹函数的区间内，有

$E[u(x)] \geqslant E[u(y)]$. 一般情况下，可证，当 $a, b \geqslant 0$ 时，$E[x] \geqslant E[y]$, $\mathrm{Var}[x] + (E[x])^2 \leqslant \mathrm{Var}[y] + (E[y])^2$, 可导出 $E[u(x)] \geqslant E[u(y)]$.

参考文献

[**1**] Allais, M., 1953, Le comportement de l'homme rationnel devant le risque，critique des postulates et axioms de l'école americaine，*Econometrica*, **21:** 503–546.

[**2**] Arrow, K., 1970, Essays in the Theory of Risk–Bearing, Amsterdam，North–Holland.

[**3**] Bernoulli, D., 1954, Exposition of a new theory on the measurement of risk, *Econometrica*, **22:** 23–36.

[**4**] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N. J..

[**5**] Duran, D., 1957, Growth stocks and the Petersburg Paradox, *Journal of Finance*, **12:** 348–363.

[**6**] Fishburn, P. C., 1988, Nolinear Preference and Utility Theory, The John Hopkins University Press, Baltimore and London.

[**7**] Friedman, M., and L. J. Savage, 1952, The expected-utility hypothesis and the measurability of utility, *Journal of Political Economy*, **60:** 463–474 (中译文见：弗里德曼文萃 (上册)，首都经济贸易大学出版社，2001，北京，196–219)。

[**8**] Hanoch, G., & H. Levy, 1969, The efficiency analysis of choices involving risk, *Review of Economic Studies*, **36:** 335–346.

[**9**] Herstein, I. N. & J. Milnor, 1953, An axiomatic approach to measurable utility, *Econometrica*, **21:** 291–297.

[**10**] Huang, Chi-fu, & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall，Englewood Cliffs，N. J..

[**11**] Ingersoll，J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers，Maryland.

[12] Jarrow，R. A., 1988，Finance Theory，Prentice-Hall，Englewood Cliffs，N. J..

[13] Kahneman, D., & A. Tversky, 1979, Prospect theory: an analysis of decision under risk, *Econometrica*, **47:** 263–291.

[14] Kahneman, D., & M. W. Riepe, 1998, Aspects of investor psychology: Beliefs, preferences, and biase investment advisors should know about, *Journal of Portfolio Management*, **24:4**, 52–65.

[15] Kreps, D. M., 1990, A Course in Microeconomic Theory, Princeton University Press, Princeton, N. J..

[16] Pratt, J., 1964，Risk aversion in the small and in the large with Applications, *Econometrica*, **32:** 143–154.

[17] Raiffa, H., Decision Analysis: Introductry Lectures on Choice under Uncertainty, Reading, MA: Addison–Wesley.

[18] Rothschild, M., & J. E. Stiglitz, 1970, Increasing risk: I. A definition, *Journal of Economic Theory*, **2**: 225–243.

[19] Rothschild, M., & J. E. Stiglitz, 1971, Increasing risk: II. Its economic consequences, *Journal of Economic Theory*, **3**: 66–84.

[20] Savege, L. J., 1954, Foundations of Statistics, Dover, New York.

[21] Tobin, J., 1958, Liquidity preference as bihavior toward risk, *Rev. Econ. Studies*, **25:** 65–86.

[22] Tversky, A. & D. Kahneman, 1981, The framing of decisions and the psychology of choice, *Science*, **211:** 453–458.

[23] Varian, H., 1992, Microeconomic Analysis, 3rd ed., W. W. Norton, New York (中译本：微观经济学高级教程，1997，周洪、李勇等译，经济科学出版社，北京)。

[24] von Neumann, J., and O. Morgenstern, 1944, Theory of Games and Economic Behavior, Princeton University Press, Princeton, New York (中译本：竞赛论与经济行为，1963，王建华、顾玮琳译，科学出版社，北京)。

[25] 史树中，1990，数学与经济，湖南教育出版社。

七、一般经济均衡与资产定价

本讲要求：对一般经济均衡的理论框架有较具体的理解。了解纯交换经济的一般经济均衡的存在定理以及它与金融市场均衡之间的关系。对不完全金融市场的一般均衡理论有明确的概念。了解 CAPM 和 APM 怎样从一般经济均衡的框架得到。

数学预备知识：线性代数。初等概率论 (随机变量的数学期望、方差、协方差等)。初等微分学（求导数，条件极值的必要条件)。

内容：

- 7.1 纯交换经济的数学表达。
- 7.2 纯交换经济的一般经济均衡的存在定理。
- 7.3 金融市场的一般均衡的存在。
- 7.4 CAPM 的均衡定价讨论。
- 7.5 APT 的均衡定价讨论。

有了期望效用函数以后，我们就可从投资者追求个人效用最大的角度来研究不确定环境下的证券市场中的最优投资组合等问题。然后，再从全市场中的投资者的投资－消费行为达到一般经济均衡，而得到全市场的资产的定价。这就为金融经济学注入了真正的经济学内容。也就是说，它以一个完整的经济模型来回答金融资产是如何定价的，而不再像线性定价法则那样，只有一个抽象空洞的框架。当然，正如我们在“代引言”所说，这样做只是一种“宏观规范”。对于商 (管理) 学院的学生或者对于从事金融实务的工作者来说，这可能并不一定意味着有可能提供金融资产实际定价的更好的方法，而是使人对金融资产定价问题的经济学有更全面深入的理解。

为讨论这样的问题，我们先回顾一下经典的一般经济均衡理论。

7.1　经典的 Arrow–Debreu 模型中把经济活动者分为两类：消费者和生产者。对我们以下的讨论来说，我们不必顾及生产者。于是模型中只剩下消费者。通常假定这些消费者的手头都有若干商品资源。由于经济中没有生产，从而这一经济中的经济活动只剩下消费者们相互之间的商品交换。这样的经济称为**纯交换经济**。在纯交换经济中的一般经济均衡的存在问题就是指其中是否存在一个一般经济均衡价格体系，使得每一个消费者只需要根据这一价格体系来卖掉手中的商品得到一份钱，再用这份钱去买他需要的商品，使得他的消费效用最大。

现在我们使这一问题数学形式化。假定经济 $\mathcal{E}$ 中有 L 种商品，I 个消费者。L 种商品的量可以用 $\mathbb{R}^L$ 中的 L 维向量 $z=(z_1,\ldots,z_L)$ 来表示，

其中每一个分量代表每一种商品的量；而每个消费者 i 则由两个因素来刻划：一是他所持有的资源商品 $\omega^i \in \mathbb{R}^L_+$, 它表示消费者 i 所持有的各种商品的量；另一是他的效用函数 u^i, 它用来刻划消费者的消费行为。这样，这个经济 $\mathcal{E}$ 也可以更加明确地表示为 $\mathcal{E}(u^1,\ldots,u^I;\omega^1,\ldots,\omega^I)=\mathcal{E}(u,\omega)$.

假定市场上已经有一个价格体系 $p=(p_1,\ldots,p_L)\in\mathbb{R}^L_+$, 它的各个分量代表各种商品的价格，并且暂且允许价格可能是零。我们来考虑消费者 i 是如何来进行他的消费决策的。这时，一个商品向量 z 的价值就是

$$p\cdot z=p_1z_1+\cdots+p_Lz_L.$$

而消费者 i 可能动用的资金也就是

$$p\cdot\omega^i=p_1\omega^i_1+\cdots+p_L\omega^i_L.$$

因此，消费者 i 可能消费的商品 z 一定要满足

$$p\cdot z\leqslant p\cdot\omega^i.$$

由于我们一般都假定效用函数满足前一讲中提出的强单调假设 (当然，目前的效用函数的定义域与前一讲中的有所不同)，故消费者为达到其最优消费，总是要把钱花光，即对他的最优消费问题，只需考虑收支平衡的情形。因此，令

$$B(\omega^i,p)=\left\{z\in\mathbb{R}^L_+\,|\,p\cdot(z-\omega^i)=0\right\}.$$

那么消费者面临的就是下列决策问题：

$$\left\{\begin{array}{ll}\max & u^i(z)\\ \text{s.t.} & z\in B(\omega^i;p)\end{array}\right. \tag{1}$$

这一问题不一定有解。下列定理在数学上是资产定价第二基本定理的简化情形，它指出，在效用函数的强单调假设下，问题有解的充分必要条件是不能有非正的商品价格。

定理 7.1 在强单调假设下，问题 (1) 有解的充分必要条件为 $p\in\mathbb{R}^L_{++}$, 这里 $\mathbb{R}^L_{++}$ 表示所有分量都为正的 L 维向量全体。

证明 如果有一种商品的价格 $p_l=0$, 那么令

$$v^l=(0,\ldots,\underbrace{1}_{\text{第 } l \text{ 个}},\ldots,0)\in\mathbb{R}^L,$$

对于任何 $z\in B(\omega^i,p)$ 总有 $z+v^l\in B(\omega^i,p)$. 由强单调假设， $u^i(z+v^l)>u^i(z)$. 因此，u^i 在 $B(\omega^i,p)$ 中不可能有最大值。

反之，如果 $p \in \mathbb{R}^L_{++}$, 那么对于任何 $z \in B(\omega^i, p)$, 显然有

$$0 \leqslant p_l z_l \leqslant p \cdot \omega^i, \qquad l = 1, \ldots, L.$$

从而 $z_l \in [0, p \cdot \omega^i / p_l]$. 这说明，$B(\omega^i, p)$ 是一个有界集，再加上它是个闭集以及 u^i 是连续函数，问题 (1) 一定有解。 □

把这条定理与资产定价第二基本定理 5.5 相比较，也可使我们对无套利假设有进一步理解。这条定理是说，如果有的商品不要钱，而不管什么商品，对消费者来说，总是越多越好，那么消费效用最大化问题就不可能有解。与资产定价第二基本定理相比较，就可以看到，无套利假设相当于没有不要钱的商品。

还有一点需要注意的是，在问题 (1) 中的价格 p 出现在一个线性齐次等式中。这意味着，如果所有商品的价格都上涨或下跌同样的倍数，那么它对原来的问题不会有任何影响。这就是说，重要的仅仅是各种商品价格之间的相互比较，而不是它们的绝对值。因此，我们总可通过某种方式来使价格规范化。常用的方式有三种。一种是假定所有价格之和为 1, 这种方式比较自然，但是从数学上来说，所有可能的价格集合形成一个 $L-1$ 维的单纯形[1]，有时会带来数学上的不方便；一种是假定所有价格的平方和为 1，这时所有可能的价格形成一个球面的一部分，尽管在现实世界很难想象这样来处理价格，但这样做在数学上有其方便之处；还有一种方式是假定第一种商品的价格为 1，历史上这是 Walras 首先这样做的，他认为这是把货币当作第一种商品，而货币的价格总是 1。我们在下面暂且对价格不作规范化假定，而在问题求解中，看哪一种规范方式方便就用哪一种。

7.2　现在我们来考虑经济 $\mathcal{E}(u, \omega)$ 中的一般经济均衡问题。这一问题可表述为：求价格体系 $p^* \in \mathbb{R}^L_{++}$ 和 $z^* = (z^{*1}, \ldots, z^{*I}) \in (\mathbb{R}^L_+)^I = \mathbb{R}^{LI}_+$, 使得

i) 对于任何 $i = 1, \ldots, I$, 对于任何 $z \in B(\omega^i, p^*)$, $u^i(z) \leqslant u^i(z^{*i})$;

ii) $z^{*1} + \cdots + z^{*I} = \omega^1 + \cdots + \omega^I$.

满足这样条件的 $(z^*, p^*) \in \mathbb{R}^{LI}_+ \times \mathbb{R}^L_+$ 称为经济 $\mathcal{E}(u, \omega)$ 中的**一般均衡** (General Equiribrium, GE)。正如我们曾经在“代引言”中介绍过的，这种类型的一般均衡的存在问题，人们经过大半个世纪才搞得比较清楚。

---

[1] 在 n 维向量空间中，$n+1$ 个不在同一个 $n-1$ 维超平面 (线性子空间的平移) 上的点所形成的最小凸集称为 n 维单纯形。例如，二维空间中的三个不共线的点所形成的单纯形就是这三个点形成的三角形。

为使一般均衡存在，我们还需要对效用函数作些假设。下面是其中之一。

严格拟凹性假设 $u\colon \mathbb{R}_+^L \to \mathbb{R}$ 是严格拟凹函数，即对于任何 $z^1, z^2 \in \mathbb{R}_+^L$ 和任何 $\lambda \in (0,1)$, $u((1-\lambda)z^1 + \lambda z^2) > \min\{u(z^1), u(z^2)\}$.

严格凹函数一定是严格拟凹函数。因此，严格拟凹函数类包括所有“边际效用严格递减”的效用函数。反之，拟凹函数不一定是凹函数。例如，单变量函数中的严格单调函数一定是严格拟凹函数，但不一定是严格凹函数。不过在多变量函数情形下，严格单调函数不一定是严格拟凹函数。否则我们就只需要强单调假设，而不再需要这条严格拟凹假设。

下列定理是 Arrow–Debreu 一般经济均衡存在定理的特殊情形。

定理 7.2 如果经济 $\mathcal{E}(u,\omega)$ 中的效用函数都满足强单调假设和严格拟凹假设，并且 $\omega^1 + \cdots + \omega^I \in \mathbb{R}_{++}^L$, 那么 $\mathcal{E}(u,\omega)$ 中存在一般均衡。

证明概要 严格完整地证明这条定理超出了本讲义的范围。但是我们可以简述一下它的证明思路。

首先定义对于每个消费者 i 关于价格 p 的需求函数 $f^i\colon \mathbb{R}_{++}^L \to \mathbb{R}^L$, 其中对于任何价格 $p \in \mathbb{R}_{++}^L$, $f^i(p)$ 就定义为问题 (1) 的解。利用强单调假设，可知这样的解是存在的；而利用严格拟凹假设，不难证明这样的解是唯一的。因此，这样定义需求函数是可能的。

其次定义全经济的 (**总量**) **超** (**过**) **需** (**求**) **函数** (aggregate excess demand function) $Z\colon \mathbb{R}_{++}^L \to \mathbb{R}^L$, 它定义为

$$Z(p) = \sum_{i=1}^{I} (f^i(p) - \omega^i).$$

其含义是总需求与总供给之差。于是一般均衡的存在问题归结为这个超需函数有没有零点。

肯定它有零点的一种方法是先把价格规范化为

$$p \in S = \{p \in \mathbb{R}_+^L \,|\, p_1 + \cdots + p_L = 1\}.$$

这里价格 p 不再要求其分量都为正。同时，把 Z 连续延拓到 S 上。简单延拓是做不到的，因为当某一商品的价格趋于零时，正如我们在定理 7.1 中所指出的，该商品的总需求会趋于无限。为消除这一困难，我们可以对问题 (1) 再加限制。例如，要求每种商品 z_l 的需求不超过该商品的总供给。于是每个消费者的需求函数 f^i 对 S 中的 p 都有定义，并且是一个连续映射。从而 Z 也是 S 上的连续映射。问题归结为一个连

续映射 $Z\colon S\to\mathbb{R}^L$ 是否有零点。有一条为证明这一结果而被人发现的 Debreu–Gale–Nikaïdo 定理指出，如果对于任何 $p\in S$, Z 满足 $p\cdot Z(p)\leqslant 0$ (它称为广义 Walras 法则，其含义是“总支出不能超过总收入”)，那么存在 $p^*\in S$, 使得 $Z(p^*)\leqslant 0$ (这是一个 L 维向量不等式，其含义是每种商品的需求都不超过供给)。这条定理与著名的 Brouwer 不动点定理 (S 到 S 的连续映射有不动点) 是等价的。因此，其证明相当困难。然后，再根据强单调假设可指出，这个 p^* 代表的价格体系就是一般均衡价格体系。事实上，由于每个消费者都不可能让供给超过需求，最终结果一定是供需相等。

另一种肯定它有零点的方法是先把价格规范化

$$p\in P=\{p\in\mathbb{R}^L_{++}\,|\,p_1^2+\cdots+p_L^2=1\},$$

再对效用函数加上光滑性假设 (函数无限次可微)。于是问题归结为求一个光滑映射 $Z\colon P\to\mathbb{R}^L$ 是否有零点的问题。这个 Z 可以理解为带边光滑流形 P 上的向量场，因为它满足 $p\cdot Z(p)=0$ (它称为 Walras 法则，其含义为“总支出等于总收入”。这个等式表示两个向量垂直，因此 $Z(p)$ 在球面的点 p 上与球面相切。) 。利用 p 趋向于边界 (某些商品的价格趋向于零) 时，超需函数中的某些商品的需求趋向于无限，可以指出，这样的向量场在接近边界时，是“外向”向量场。于是根据它可以构造一个与它有同样零点的光滑流形 P 上的“内向”光滑向量场。利用所谓 Poincaré–Hopf 定理，可指出这样的向量场有零点。 □

与一般均衡紧密相连的还有一个问题是所谓“资源最优配置”问题。在上述模型中，任何满足 $z^1+\cdots+z^I\leqslant\omega^1+\cdots+\omega^I$ 的 L 维商品组 $z=(z^1,\ldots,z^I)\in\mathbb{R}^{LI}_+$ 称为 $\mathcal{E}(u,\omega)$ 中的一个 (**可行资源**) **配置** (allocation)。所谓最优配置是指在 Pareto 意义下的最优，即在不损害别的消费者利益的前提下，每个消费者不可能再增加自己的效用。它的数学表达如下： $z^*=(z^{*1},\ldots,z^{*I})\in\mathbb{R}^{LI}_+$ 称为是 $\mathcal{E}(u,\omega)$ 中的 **Pareto 最优配置**，是指不存在另一个配置 $z=(z^1,\ldots,z^I)\in\mathbb{R}^{LI}_+$, 使得下列不等式中至少有一个严格不等式：

$$u^i(z^i)\geqslant u^i(z^{*i}),\qquad i=1,\ldots,I.$$

下列定理经常被称为**福利经济学第一定理**：

定理 7.3 在定理 7.2 的条件下，如果 $(z^*,p^*)\in\mathbb{R}^{LI}_+\times\mathbb{R}^L_{++}$ 是 $\mathcal{E}(u,\omega)$ 中的一般均衡，那么 z^* 是 $\mathcal{E}(u,\omega)$ 中的 Pareto 最优配置。

证明 如果 z^* 不是 Pareto 最优配置，那么存在配置 z 满足 $u^i(z^i)\geqslant u^i(z^{*i})$, $i=1,\ldots,I$, 并且存在 j, 满足 $1\leqslant j\leqslant I$, 使得 $u^j(z^j)>u^j(z^{*j})$. 因为 u^j 在 z^* 处取其在 $B(\omega^j,p^*)$ 中的最大值，故 $z^j\notin B(\omega^j,p^*)$, 即

$p^* \cdot z^j > p^* \cdot \omega^j$. 另一方面，由强单调假设，我们也有 $p^* \cdot z^i \geqslant p^* \cdot \omega^i$, $i = 1, \dots, I$. 这是因为如果对于某个 i 有 $p^* \cdot z^i < p^* \cdot \omega^i$, 我们总可找到一个非零向量 $v \in \mathbb{R}_+^L$ 使得 $p^* \cdot (z^i + v) = p^* \cdot \omega^i$, 从而 $u^i(z^i) < u^i(z^i + v) \leqslant u^i(z^{*i})$, 与假设矛盾。最后，我们得到

$$p^* \cdot (z^1 + \cdots + z^I) > p^* \cdot (\omega^1 + \cdots + \omega^I).$$

这与 $z = (z^1, \dots, z^I)$ 是配置矛盾。 □

这条“福利经济学第一基本定理”的“逆定理”也成立，即，对每一个 Pareto 最优配置都存在某种价格体系，使得它们形成一般均衡。它称为“福利经济学第二基本定理”。其证明需要利用凸集分离定理。这里不再细述。

7.3　现在我们来讨论有关最优证券组合选择的一般均衡存在问题。我们假定目前有 I 个投资者，经济中没有商品，而是如前一讲中所讨论的未定市场那样，有当前和未来两个时刻，未来有 S 种状态，市场中有 K 种证券，如此等等。每个投资者 i 要考虑的问题就是前一讲中考虑的问题

$$\begin{cases} \max & u^i(z) \\ \text{s.t.} & z \in B(\omega^i; x^0, x^{\mathbf{1}}) \end{cases} \tag{2}$$

其中

$$B(\omega^i; x^0, x^{\mathbf{1}}) = \left\{ z \in \mathbb{R}_+^{S+1} \;\middle|\; \begin{array}{c} z^0 = \omega^{i0} - \theta \cdot x^0, \\ z^{\mathbf{1}} = \omega^{i\mathbf{1}} + \theta \cdot x^{\mathbf{1}}, \\ \theta \in \mathbb{R}^K. \end{array} \right\}.$$

这就是说，目前我们考虑的投资者 i 是用一个衡量他的当前和未来消费的效用函数 u^i 和他所掌握的当前与未来所持有的资金 ω^i 来刻划的。我们仍然利用同样的记号，这样便于与前面的讨论比较。但是需要注意的是其经济含义已经有所改变。

目前要考虑的一般均衡问题不再是要给商品定价，而是要给 K 种证券的当前定价。其中 K 种证券的未来价格则是被认为用支付矩阵给定的。因此，我们的经济现在要表示为 $\mathcal{E}(u, \omega, x^{\mathbf{1}})$, 其中 $u = (u^1, \dots, u^I)$, $\omega = (\omega^1, \dots, \omega^I)$, $u^i \colon \mathbb{R}_+^{S+1} \to \mathbb{R}_+$, $\omega^i \in \mathbb{R}_+$, $i = 1, \dots, I$, 而 $x^{\mathbf{1}} = (x_k^s)_{k=1,\dots,K; s=1,\dots,S}$ 是 K 种证券的 S 种未来价格所形成的支付矩阵。

$\mathcal{E}(u, \omega, x^{\mathbf{1}})$ 可称为**金融市场**。在这个金融市场中，我们期望的是通过每个投资者求其最优证券组合选择来形成均衡的证券当前价格。需要注意的是，在这里，证券的当前价格有可能是负的。金融市场的一般均衡定义为 $(z^*, \theta^*, x^{*0}) \in \mathbb{R}_+^{I(S+1)} \times \mathbb{R}^{I(K)} \times \mathbb{R}^K$ 使得

i) 对于任何 $i = 1, \dots, I$, 对于任何 $z \in B(\omega^i, x^{0*}, x^{\mathbf{1}})$, $u^i(z) \leqslant u^i(z^{*i})$;

ii) $\theta^{*1} + \cdots + \theta^{*I} = 0;$

ii) $z^{*1} + \cdots + z^{*I} = \omega^1 + \cdots + \omega^I.$

这里条件 ii) 是指证券交易就在这些投资者之间进行，所有投资者所掌握的证券组合之和使市场结清。在经济含义上，它意味着这里的证券都是债券型的，而不是股票型的。但是这并不影响模型的一般性。我们也可在模型中假设还有一些投资者持有总和不为零的证券。例如其总和恒为 1，它可意味着每个投资者都掌握某公司股票的一定份额 (可能为零或负值)。对于这种情况很容易通过简单的变换在数学上归结为上述情形。具体的数学表达可参看下节。而条件 iii) 实际上是条件 ii) 的直接推论。这样的一般均衡称为**金融市场均衡** (Financial Market Equiribrium, FME)。

我们把这一模型与前面的纯交换经济模型相比较，可以看到，金融市场中的交易与纯交换经济中的交易在某种意义上是一样的。把不同状态下的钱看成不同的商品，前者相当于人们在把不同状态下的钱相互交换，以求得各自的最大效用。而证券在其中只起中介作用。

金融市场均衡的存在问题比起前节讨论的确定性商品市场的一般均衡的存在问题来，当然更加复杂。但是由于前一讲中的资产定价第二基本定理的成立，我们首先可以知道，如果金融市场均衡存在，那么在这一市场中，无套利假设必须成立。如果事先加上无套利假设，那么求证券当前价格的问题，可转化为求状态价格 (或者说求 Arrow–Debreu 证券的价格) 的问题。引入状态价格 $\lambda = (\lambda_1, \ldots, \lambda_S) \in \mathbb{R}^S_{++}$, 我们可以看到，问题 (2) 转化为

$$\begin{cases} \max & u^i(z) \\ \text{s.t.} & z \in B(\omega^i; \lambda, x^{\mathbf{1}}) \end{cases} \tag{3}$$

其中

$$B(\omega^i; \lambda, x^{\mathbf{1}}) = \left\{ z \in \mathbb{R}^{S+1}_+ \;\middle|\; \begin{array}{c} z^0 - \omega^{i0} + \lambda_1(z^1 - \omega^{i1}) + \cdots + \lambda_S(z^S - \omega^{iS}) = 0, \\ z^{\mathbf{1}} - \omega^{i\mathbf{1}} = \theta \cdot x^{\mathbf{1}}, \quad \theta \in \mathbb{R}^K. \end{array} \right\}.$$

其相应的一般均衡也可类似定义。

现在我们分两种情形来讨论。一种是完全市场情形；另一种是不完全市场情形。在完全市场情形下 (这时 $K \geqslant S$)，我们可以发现上面的约束集 $B(\omega^i; \lambda, x^{\mathbf{1}})$ 中的 $x^{\mathbf{1}}$ 不起作用，因为对于任何 S 维向量 (未定权益) $z^{\mathbf{1}} - \omega^{i\mathbf{1}}$, 我们总能找到适当的组合 θ, 使得 θ 的未来价值与它一样，即 $z^{\mathbf{1}} - \omega^{i\mathbf{1}} = \theta \cdot x^{\mathbf{1}}$. 这样，$B(\omega^i, \lambda, x^{\mathbf{1}}) = B(\omega^i, \lambda)$. 把它与上一小节中讨论的一般均衡相比较，我们可以看到，经过这样的转化以后，现

在的问题在形式上与前面的问题几乎完全一样，仅有的区别除了空间的维数不同外，相当于原来的价格向量 p 的状态价格向量 λ 维数少了一维。如果认为当前状态的“状态价格” $\lambda_0 = 1$, 并把它与原来的 λ 合在一起形成一个价格向量，那么它无非是 Walras 的价格规范形式，其他方面与前面就完全一致了。综上所述，我们就可得到下列定理：

定理 7.4 如果金融市场 $\mathcal{E}(u, \omega, x^{\mathbf{1}})$ 是完全的，其中的效用函数都满足强单调假设和严格拟凹假设，并且 $\omega^1 + \cdots + \omega^I \in \mathbb{R}^{S+1}_{++}$, 那么 $\mathcal{E}(u, \omega, x^{\mathbf{1}})$ 中存在均衡，并且这一均衡是 Pareto 最优配置。

这条定理更加具体地说明了我们前面提到的意思：在完全金融市场中，金融市场均衡与纯交换经济的一般均衡在原理上是完全一样的。当年 Arrow 和 Debreu 也就是这样认为带不确定性的金融市场的均衡问题可归结为确定性的商品市场的均衡问题。完全金融市场均衡经常称为**未定市场均衡** (Contingent Market Equilibrium, CME)，其中“未定市场 (contingent market)”这个词正如我们已经提到的，并没有什么字面意义。

然而，这恰恰是说明完全市场没有反映不确定性的本质。而不完全市场的均衡存在问题要复杂得多。在完全市场情形下，证券的未来价值支付矩阵 $x^{\mathbf{1}}$ 其实在求均衡时没有起多大作用。除了开始由它来确定市场是完全的以外，在求均衡的过程中，它不再起作用。最后为求出均衡的证券组合选择时，才再次起用它。但是在不完全市场的情形下，它对求均衡的过程起很大的约束作用，以至长期来人们不知怎样来处理它。金融市场模型是 1972 年由 Radner [14] 首先提出的，并且他在对卖空有上界 (不能无限制地卖空) 的条件下，证明了均衡的存在。但是 1975 年 Hart [7] 用反例指出，不完全金融市场的一般均衡可能不存在[2]。这使得这方面的研究长期停滞不前。 1985 年，Duffie 和 Schafer [5], [6] 用相当复杂的数学方法指出，对于“极大多数”的不完全市场，均衡还是存在的。这里的“极大多数”是指对于所有可能的初始持有 ω 与所有可能的证券未来价值支付矩阵 $x^{\mathbf{1}}$ 所形成的集合而言的 (均衡不存在的集合测度为零)。他们同时还证明了，不完全市场的“极大多数”均衡都是 Pareto 无效的，不能达到“资源最优配置”。不完全市场的均衡不能达到 Pareto 最优配置，是人们曾经想到的，因为市场不完全，风险不能完全规避，仅仅通过交易似乎不能达到最优。但是把它们明确表达为这样的定理，那还是第一次。这些研究结果都值得人们深思。

---

[2] 可以证明，二期“单商品” (只有“钱”是唯一的一种商品) 金融市场的的一般均衡 (不管市场是否完全) 总是存在的。但是对于二期多商品 (除了“钱”以外，还有别的商品，从而模型中可包含“期货合约”) 不完全金融市场，或者三期以上的单商品市场，一般均衡就可能不存在。 Hart 的反例是对于二期二商品市场构造的。这一反例也很容易改造为对于三期单商品市场的反例。参看 Magill–Quinzii [11]。

不完全市场的一般均衡 (General Equilibrium of Incomplete Markets, GEI) 理论在近十几年来有很大发展。Magill 和 Quinzii 的 [11] 是近年来这方面出现的内容非常丰富的专著。

7.4 下面两节中，我们将对 CAPM 和 APT 用一般均衡理论来讨论。这样的讨论可以在未来状态有限的框架下进行。办法是对每一状态赋以客观概率，以至每一未来价格都可看作有分布的随机变量。事实上，在 Magill–Quinzii [11] 中就是这样来讨论 CAPM 的，其中期望效用函数采用 6.5 中所提到的形式，即 $u(z^0, z^1, \ldots, z^S) = v(z^0) + \delta\sum_{s=1}^{S} q_s v(z^s)$, 而 v 则采用 HARA 类函数。由此可以得到同样形式的 CAPM。

然而，这样的讨论离开已有的 CAPM 的叙述较远。我们在下面将讨论证券未来价格是一般的随机变量的情形。这样的讨论在数学上有一定困难，但是如果我们避开均衡存在问题，将得到更接近经典的 CAPM (以及 APT) 的结果。

我们仍假定只有当前和未来两个时刻，当前是确定的，但未来是不确定的。假定市场中有 n 种基本风险证券，其未来价格是 n 个随机变量 $x_1, x_2, \ldots, x_n$; 这里假设它们间线性无关，从而其协方差矩阵和收益率的协方差矩阵正定。此外，第 0 种证券是无风险证券，其未来价格 x_0 是确定量。我们仍假定线性定价法则成立，以至存在某个线性定价函数 p, 使得这 n 种证券的当前价格为 $p(x_0), p(x_1), p(x_2), \ldots, p(x_n)$, 且假定它们都为正。这 $n+1$ 种证券的投资组合可用 $n+1$ 维向量 $\theta = (\theta_0, \theta_1, \ldots, \theta_n)$ 来表示。那么投资组合的当前价格为

$$y = \theta_0 p(x_0) + \theta_1 p(x_1) + \cdots + \theta_n p(x_n); \tag{4}$$

而投资组合的未来价格为

$$y = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n. \tag{5}$$

类似前面讨论的情形，又假设有某个投资者，他在当前持有 (确定的) 资金 ω^0, 未来则持有 (随机的) 的资金 ω^1. 他的消费效用是用某个作为当前 (确定的) 消费 z^0 和未来 (随机的) 消费 z^1 的函数的期望效用函数 $u(z^0, z^1)$ 来衡量的。这样他面临的最优证券选择问题如下：求当前消费 z^0 和投资组合 θ, 使得 $E[u(z^0, z^1)]$ 达到最大。这里 u 只对确定量来定义，从而它的期望值就是对随机变量的期望效用。更具体地说，问题可表述为：求证券组合 θ, 使得

$$\left\{\begin{array}{ll} \max & E[u(z^0, z^1)] \\ \text{s.t.} & z^0 = \omega^0 - \sum_{k=0}^{n} \theta_k p(x_k), \\ & z^1 = \omega^1 + \sum_{k=0}^{n} \theta_k x_k. \end{array}\right. \tag{6}$$

这里第一个等式是确定量的等式，第二个等式是随机量的等式。

这一问题是否可解的讨论要比前面有限状态的情形复杂得多。虽然对这里的函数 u 我们也可以加上强单调之类的假设，但是由于这里还涉及数学期望运算，很难得到以前那样的简洁定理。不过我们可对一些具体的期望效用函数来进行讨论。

以下我们的讨论目标是：由此出发导出资本资产定价模型 (CAPM)。为此，我们先假定投资者在当前还掌握一个证券组合 θ^0，令 $\theta = \theta^1 - \theta^0$, 并用 ω^1 取代 $\omega^1 - \sum_{k=0}^n \theta_k^0 x_k$，那么上述问题转化为下列问题：求证券组合 θ^1, 使得

$$\begin{cases} \max & E[u(z^0, z^1)] \\ \text{s.t.} & z^0 = \omega^0 - \sum_{k=0}^n (\theta_k^1 - \theta_k^0) p(x_k), \\ & z^1 = \omega^1 + \sum_{k=0}^n \theta_k^1 x_k. \end{cases} \tag{7}$$

这样提出的问题在经济含义上与以前的问题有所不同。它相当于金融市场中有股票型的证券。投资者原来有证券组合 θ^0, 在当前把它套现，并再购买新的组合 θ^1. 在以前的金融市场均衡模型中，我们要求市场中所有证券组合之和 $\sum_{i=1}^I \theta^i = 0$. 这样的证券将更多地意味着债券或投资者之间的金融合约。而现在要讨论的均衡模型中，我们对均衡将要求 $\sum_{i=1}^I (\theta^{0i} - \theta^{1i}) = 0$, 但 $\sum_{i=1}^I \theta^{0i} = \sum_{i=1}^I \theta^{1i} \gg 0$ ($\gg$ 表示向量的所有分量为正)，它们将代表全市场的证券组合。虽然在数学上实质上没有本质改变，但是经济含义上已经有了变化。

我们假定 $E[u(z^0, z^1)] = v(z^0, E[z^1], \text{Var}[z^1])$, 即效用函数只与当前消费以及未来消费的均值和方差有关。与此相联系的当然是函数的未来消费自变量应该与某个“未定权益 Hilbert 空间”有关。在经济学的讨论中，通常总假定期望效用函数光滑、递增、严格凹 (风险厌恶) 等。Markowitz 在他的证券组合选择理论中，开始时没有使用期望效用函数，而仅以组合的收益率的均值和方差来衡量组合收益的优劣，即只提出所谓“均值－方差准则”。对于投资者来说，如果要用效用函数来决策，那么他可以采用一个只与收益率的均值和方差有关的函数来作为效用函数。这与传统的期望效用函数有所不同。Markowitz 和 Tobin 为了说明这样的处理方法与传统的处理方法不矛盾，曾经指出，这相当于期望效用函数是二次凹函数 $u(x) = ax - bx^2, a, b > 0$. 事实上，如果 u 如上定义 (请注意，这里的效用函数是对未来消费的价值来定义的，而不是对证券组合的收益率定义的)，那么

$$E[u(x)] = aE[x] - bE[x^2] = aE[x] - b(\text{Var}[x] + E^2[x]).$$

反之，如果假定 u 有 Taylor 展开，那么也容易指出， $E[u(x)]$ 只与 x 的均值和方差有关必定是二次函数。不过在这种情况下，如果还要求它

关于 x 是递增凹函数，那么对 x 的变化范围将有所限制。此外，还可看到，如果所涉及的随机变量都是正态分布的随机变量，那么这一要求自然满足。注意到这些以后，我们就可认为，目前的假定在一定程度上是合理的。然而，正如我们在上一讲讨论随机占优问题时已经提到，这里要引起注意的是：Markowitz 的均值－方差准则从根本上来说还是与期望效用函数理论不完全一致的。它至多是上述假定下的一类特殊函数的随机占优的充分条件，但并非必要条件。这实际上说明，Markowitz 的均值－方差准则不可能用 von Neumann–Morgenstern 期望效用函数公理来刻划。

在这样的假定下，如果再假定 ω^1 是定常量, 那么问题 (7) 可以表达得更明确。这是因为 $\mu_{z^1}=E[z^1]$ 和 $\sigma^2_{z^1}=\mathrm{Var}[z^1]$ 在目前都可算出：

$$\mu_{z^1}=\sum_{k=0}^{n}\theta_k^1E[x_k],$$

$$\sigma^2_{z^1}=\sum_{j,k=1}^{n}\theta_j^1\theta_k^1\mathrm{Cov}[x_j,x_k].$$

这样，问题 (6) 变为：求证券组合 θ^1, 使得

$$\begin{cases}\max & v(z^0,\mu_{z^1},\sigma^2_{z^1})\\ \text{s.t.} & z^0=\omega^0+\sum_{k=0}^{n}(\theta_k^0-\theta_k^1)p(x_k),\\ & \mu_{z^1}=\sum_{k=0}^{n}\theta_k^1E[x_k],\\ & \sigma^2_{z^1}=\sum_{j,k=1}^{n}\theta_j^1\theta_k^1\mathrm{Cov}[x_j,x_k].\end{cases}\tag{8}$$

下面，我们为简化记号，记 $v=v(x,y,z)$, 它对三个变量的偏导数分别记为 v_x,v_y,v_z. 尤其是我们要求

$$v_x>0,\quad v_y>0,\quad v_z<0.\tag{9}$$

它们分别意味着：“当前消费价值越大越好”；“未来消费价值越大越好”；“未来消费价值的风险越小越好”。

定理 7.5 问题 (8) 的解 $(\bar\theta^1)$ 满足下列方程:

$$E[r_j]-r_0+2\frac{v_z(\bar z^0,\mu_{\bar z^1},\sigma^2_{\bar z^1})}{v_y(\bar z^0,\mu_{\bar z^1},\sigma^2_{\bar z^1})}\sum_{k=1}^{n}\bar\theta_k^1p(x_k)\mathrm{Cov}[r_k,r_j]=0,\quad j=1,\ldots,n.\tag{10}$$

$$r_0=\frac{v_x(\bar z^0,\mu_{\bar z^1},\sigma^2_{\bar z^1})}{v_y(\bar z^0,\mu_{\bar z^1},\sigma^2_{\bar z^1})},\tag{11}$$

这里 $\bar z^0$, $\bar z^1$ 为对应 $\bar\theta^1$ 的 z^0, z^1; $r_0,r_1,\ldots,r_n$ 分别为各证券的收益率。

证明　证明可用 Lagrange 乘子法。问题 (8) 的 Lagrange 函数为

$$L = v\left(z^0, \sum_{k=0}^{n}\theta_k^1 E[x_k], \sum_{j,k=1}^{n}\theta_j^1\theta_k^1 \mathrm{Cov}[x_j, x_k]\right)$$

$$+\lambda\left(\omega^0 + \sum_{k=0}^{n}\theta_k^0 p(x_k) - z^0 - \sum_{k=0}^{n}\theta_k^1 p(x_k)\right).$$

因此，其解 $\bar{\theta}$ 和相应的 $(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2)$ 满足下列条件：

$$\frac{\partial L}{\partial z^0} = 0, \quad \frac{\partial L}{\partial \theta_j^1} = 0,\ j = 0, 1, \ldots, n; \quad \frac{\partial L}{\partial \lambda} = 0.$$

从而由前两式可得

$$v_x(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2) = \lambda, \tag{12}$$

$$v_y(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2)E[x_j] + 2v_z(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2)\sum_{k=1}^{n}\bar{\theta}_k^1 \mathrm{Cov}[x_j, x_k] - \lambda p(x_j) = 0, \tag{13}$$

$$j = 0, 1, \ldots, n.$$

当 $j = 0$ 时，(13) 可导致

$$v_y(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2)x_0/p(x_0) = \lambda.$$

再注意到 $x_0/p(x_0) = r_0$ 和 (12), 即得 (11). 当 $j \neq 0$ 时，同样可注意到 $E[x_j]/p(x_j) = E[r_j]$, $p(x_j)p(x_k)\mathrm{Cov}[r_j, r_k] = \mathrm{Cov}[x_j, x_k]$ 以及 λ 的上述表达式，由 (13) 可导得 (10). □

以上说的是一个投资者的最优证券组合选择。如果现在市场上有 I 个这样的投资者，即他们各有各的原有资金 ω^{0i}、证券组合 θ^{0i} 和效用函数 v^i, 以至都面临这样的证券选择问题：

$$\left\{\begin{array}{ll} \max & v^i(z^0, \mu_{z^1}, \sigma_{z^1}^2) \\ \text{s.t.} & z^0 = \omega^{0i} + \sum_{k=0}^{n}(\theta_k^{0i} - \theta_k^1)p(x_k), \\ & \mu_{z^1} = \sum_{k=0}^{n}\theta_k^1 E[x_k], \\ & \sigma_{z^1}^2 = \sum_{j,k=1}^{n}\theta_j^1\theta_k^1 \mathrm{Cov}[x_j, x_k]. \end{array}\right. \quad i = 1, \ldots, I. \tag{14}$$

并且都各自作出其最优组合选择 $\bar{\theta}^{1i}$, 那么我们就可考虑对这样的证券市场的均衡存在问题。如果适当的证券当前价格 $p(x_0), p(x_1), \ldots, p(x_n)$ 使得市场上证券刚好达到供需平衡时，即

$$\sum_{i=1}^{I}\theta_k^{0i} = \sum_{i=1}^{I}\bar{\theta}_k^{1i} \gg 0, \qquad k = 0, 1, \ldots, n, \tag{15}$$

那么称该市场达到**均衡**。在目前这样的情况下，均衡是否存在是个更加复杂的问题。这里我们将简单地**假定市场均衡存在**。

下一定理说明市场均衡的一些特点，其中最主要的一点是资本资产定价模型成立。

定理 7.6 在上述模型中，如果对于定价 $p(x_0)$, $p(x_1)$, ..., $p(x_n)$ 来说, $\bar{\theta}^{11}$, $\bar{\theta}^{12}$, ..., $\bar{\theta}^{1I}$, 形成市场均衡，那么

i) $\sum_{i=1}^{I} \omega^{i0} = \sum_{i=1}^{I} \bar{z}^{i0}$; (所有投资者的最优当前消费之和等于他们手头有的资金之和，即，总体来说，当前消费并未动用证券市场中的资金。)

ii) $\bar{\theta}_k^{1i} > 0, k = 1, \ldots, n, i = 1, 2, \ldots, I$; (每个投资者的最优证券投资不需要卖空，并且每种证券都要买；以下甚至还证明了每个投资者的对收益率来说的风险证券投资组合都是一样的。)

iii) 设 $r_{\tilde{\theta}^{1i}}$ 为第 i 个投资者的最优组合 $\bar{\theta}^{1i}$ 中的风险证券组合 $\tilde{\theta}^{1I} = (0, \bar{\theta}_1^{1i}, \bar{\theta}_2^{1i}, \ldots, \bar{\theta}_n^{1i})$ 的收益率。那么

$$E[r_j] - r_0 = \frac{\mathrm{Cov}[r_j, r_{\tilde{\theta}^{1i}}]}{\mathrm{Var}[r_{\tilde{\theta}^{1i}}]}(E[r_{\tilde{\theta}^{1i}}] - r_0), \quad j = 1, 2, \ldots, n.$$

iv) (CAPM) 设 r_m 为风险证券的市场组合 $m = (0, \sum_{i=1}^{I} \theta_1^{0i}, \sum_{i=1}^{I} \theta_2^{0i}, \ldots, \sum_{i=1}^{I} \theta_n^{0i})$ 的收益率。那么

$$E[r_j] - r_0 = \frac{\mathrm{Cov}[r_j, r_m]}{\mathrm{Var}[r_m]}(E[r_m] - r_0), \quad j = 1, 2, \ldots, n.$$

v) 设 $z^1 = \sum_{i=1}^{I} z^{1i}$ 为未来的总消费，r_{z^1} 为其相应的收益率。那么

$$E[r_j] - r_0 = \frac{\mathrm{Cov}[r_j, r_{z^1}]}{\mathrm{Var}[r_{z^1}]}(E[r_{z^1}] - r_0), \quad j = 1, 2, \ldots, n.$$

(以上是三种资本资产定价模型的表示，其形式完全一样。)

vi) (共同基金定理)

$$\sum_{k=1}^{n} \bar{\theta}_k^{1i} x_k = \left(\omega^{0i} + \sum_{k=0}^{n} \theta_k^{0i} p(x_k) - z^{0i} \right) r_m, \quad i = 1, 2, \ldots, I.$$

(最优风险证券组合可通过市场组合来实现，即市场组合可看作一种共同基金。)

证明 i) 因为

$$\omega^{0i} + \sum_{k=0}^{n} \theta_k^{0i} p(x_k) = z^{0i} + \sum_{k=0}^{n} \bar{\theta}_k^{1i} p(x_k), \quad i = 1, \ldots, I;$$

所以两端对 i 求和并移项，可得

$$\sum_{i=1}^{I} \omega^{0i} - \sum_{i=1}^{I} z^{0i} = \sum_{i=1}^{I}\sum_{k=0}^{n} p(x_k)(\bar{\theta}_k^{1i} - \theta_k^{0i}) = \sum_{k=0}^{n} p(x_k) \sum_{i=1}^{I} (\bar{\theta}_k^{1i} - \theta_k^{0i}).$$

由均衡的定义 (15)，上式为零，即 i) 成立。

ii) 由 (10), 可得

$$E[r_j] - r_0 = A^i \sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \mathrm{Cov}[r_k, r_j], \quad j = 1, \ldots, n, \tag{16}$$

其中，由 (9),

$$A^i = -2 \frac{v_z^i(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2)}{v_y^i(\bar{z}^0, \mu_{\bar{z}^1}, \sigma_{\bar{z}^1}^2)} > 0. \tag{17}$$

把 (16) 写成矩阵形式，则可得

$$E[r] - r_0 e = A^i V \bar{\theta}^{1i} p(x), \quad i = 1, \ldots, I;$$

其中 $E[r] = (E[r_1], \ldots, E[r_n])^T$, $e = (1, \ldots, 1)^T$, $\bar{\theta}^{1i} p(x) = (\bar{\theta}_1^{1i} p(x_1), \ldots, \bar{\theta}_n^{1i} p(x_n))^T$, $V = (V_{jk})_{j,k=1,\ldots,n} = (\mathrm{Cov}[r_j, r_k])_{j,k=1,\ldots,n}$. 因此，

$$A^i \bar{\theta}^{1i} p(x) = V^{-1}(E[r] - r_0 e), \quad i = 1, \ldots, I. \tag{18}$$

上式右端与 i 无关，从而可得

$$A^1 \bar{\theta}_k^{11} p(x_k) = A^2 \bar{\theta}_k^{12} p(x_k) = \cdots = A^I \bar{\theta}_k^{1I} p(x_k), \quad k = 1, \ldots, n. \tag{19}$$

由于 $A^i, p(x_k) > 0$, 故对于固定的 k, 所有 $\bar{\theta}_k^{1i}$ 同号。但是对于均衡来说，由 (15), 每种证券的总数在经济中总是正的，因此，必须所有 $\bar{\theta}_k^{1i} > 0$, 即 ii) 成立。

iii) 令

$$w_k^i = \bar{\theta}_k^{1i} p(x_k) / \sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k), \quad k = 1, \ldots, n. \tag{20}$$

那么由 (16) 可得

$$E[r_j] - r_0 = A^i \sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \sum_{k=1}^{n} w_k^i \mathrm{Cov}[r_k, r_j] = A^i \left(\sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \right) \mathrm{Cov}[r_{\tilde{\theta}^{1i}}, r_j].$$

我们只需指出，上式右端的系数就是 iii) 中的系数。事实上，由上式可得

$$\sum_{j=1}^{n} w_j^i (E[r_j] - r_0) = A^i \left(\sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \right) \sum_{j=1}^{n} w_j^i \mathrm{Cov}[r_{\tilde{\theta}^{1i}}, r_j].$$

但是

$$\sum_{j=1}^{n} w_j^i (E[r_j] - r_0) = E[r_{\tilde{\theta}^{1i}}] - r_0;$$

$$\sum_{j=1}^{n} w_j^i \mathrm{Cov}[r_{\tilde{\theta}^{1i}}, r_j] = \mathrm{Cov}[r_{\tilde{\theta}^{1i}}, r_{\tilde{\theta}^{1i}}] = \mathrm{Var}[r_{\tilde{\theta}^{1i}}].$$

由此即得

$$A^i \left(\sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \right) = (E[r_{\tilde{\theta}^{1i}}] - r_0) / \mathrm{Var}[r_{\tilde{\theta}^{1i}}]. \tag{21}$$

即 iii) 成立。

iv) 我们只需指出，对于固定的 k, 所有 w_k^i 都相等，并且都等于市场组合相应的比例系数 w_k^m. 事实上，由 (18), 我们可得

$$A^i \left(\sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \right) w^i = V^{-1}(E[r] - r_0 e), \quad i = 1, \ldots, I.$$

其中 $w^i = (w_1^i, \ldots, w_n^i)^T$. 但是由 (19), $A^i \left(\sum_{k=1}^{n} \bar{\theta}_k^{1i} p(x_k) \right)$ 与 i 无关。因此，w^i 与 i 无关，即对于固定的 k, 所有 w_k^i 都相等。另一方面，

$$\begin{aligned} w_k^m &= \frac{\sum_{i=1}^{I} \theta_k^{0i} p(x_k)}{\sum_{k=1}^{n} \sum_{i=1}^{I} \theta_k^{0i} p(x_k)} \\ &= \frac{\sum_{i=1}^{I} \theta_k^{1i} p(x_k)}{\sum_{k=1}^{n} \sum_{i=1}^{I} \theta_k^{1i} p(x_k)} \\ &= \frac{\sum_{i=1}^{I} w_k^i \sum_{k=1}^{n} \theta_k^{1i} p(x_k)}{\sum_{k=1}^{n} \sum_{i=1}^{I} \theta_k^{1i} p(x_k)} \\ &= w_k^i \frac{\sum_{i=1}^{I} \sum_{k=1}^{n} \theta_k^{1i} p(x_k)}{\sum_{k=1}^{n} \sum_{i=1}^{I} \theta_k^{1i} p(x_k)} = w_k^i. \end{aligned}$$

这样，实际上，对于任何 i, $r_{\tilde{\theta}^{1i}} = r_m$, 因而 iv) 由 iii) 可得。

v) 这是因为

$$\bar{z}^1 = \sum_{i=1}^{I} \left(\omega^{1i} + \sum_{k=0}^{n} \bar{\theta}_k^{1i} x_k \right),$$

它等于市场组合与一项无风险资产之和。由 (16) 可得

$$(A^i)^{-1}(E[r_j]-r_0)=\sum_{k=1}^{n}\bar{\theta}_k^{1i}\mathrm{Cov}[x_k,r_j],\quad j=1,\ldots,n, \tag{22}$$

其中 A^i 由 (17) 所定义。因此，

$$\begin{aligned}\left(\sum_{i=1}^{I}(A^i)^{-1}\right)(E[r_j]-r_0) &= \sum_{i=1}^{I}\sum_{k=1}^{n}\bar{\theta}_k^{1i}\mathrm{Cov}[x_k,r_j]\\ &= \mathrm{Cov}[\bar{z}^1,r_j]\\ &= p(\bar{z}^1)\mathrm{Cov}[r_{\bar{z}^1},r_j]\quad j=1,\ldots,n.\end{aligned} \tag{23}$$

另一方面，$r_{\bar{z}^1}=\sum_{k=0}^{n}w_k^{\bar{z}^1}r_k$, 其中 $w_k^{\bar{z}^1}$ 按常规定义。从而由 (23) 可得

$$\begin{aligned}\left(\sum_{i=1}^{I}(A^i)^{-1}\right)\sum_{k=0}^{n}w_k^{\bar{z}^1}(E[r_j]-r_0) &= \left(\sum_{i=1}^{I}(A^i)^{-1}\right)(E[r_{\bar{z}^1}]-r_0)\\ &= p(\bar{z}^1)\mathrm{Var}[r_{\bar{z}^1}],\quad j=1,\ldots,n.\end{aligned} \tag{24}$$

把 (23) 的两端与 (24) 的两端相除，即得 v).

vi) 事实上，

$$\sum_{k=1}^{n}\bar{\theta}_k^{1i}x_k=\left(\sum_{k=1}^{n}\bar{\theta}_k^{1i}p(x_k)\right)(r_{\tilde{\theta}^{1i}})\left(\omega^{0i}+\sum_{k=0}^{n}\theta_k^{0i}p(x_k)-z^{0i}\right)(r_{\tilde{\theta}^{1i}}),$$

$$i=1,2,\ldots,I.$$

但是 $r_{\tilde{\theta}^{1i}}=r_m$, $i=1,\ldots,I$, 故 vi) 成立。 □

以上就是用“均衡定价论”来讨论的资本资产定价模型。我们可以看到，由此得到的内容比单纯由线性定价原则出发和 Markowitz 的均值—方差分析得到的内容要丰富得多。尤其是，这里的讨论是真正有“经济学内容”的，并且无套利假设只是隐含在一般均衡存在的假定之中。但是所得到的结论中最重要的还是我们前面已经导得的通常的资本资产定价模型 (CAPM)。另一方面，我们还可看到，市场均衡的概念在上述论述中起关键作用。不过，正如我们前面已经提到的，这里的均衡是否存在还是一个问题。此外，在 Magill–Quinzii [11] 的有关 CAPM (假定未来状态有限)的讨论中指出，CAPM 均衡一定是 Pareto 最优配置。这是一个令人注目的结果。它说明，除非是在完全市场情形下， CAPM 均衡的存在是一种“稀罕”现象。

7.5 最后，我们再用“均衡定价论”来重新考察 Ross 的 APT 理论。为此，我们先来讨论一个与前面类似的最优证券组合选择问题：求当前消费 z^0 和投资组合 θ^1, 使得

$$\begin{cases} \max & E[u(z^0,z^1)] \\ \text{s.t.} & z^0=\omega^0+\sum_{k=0}^n(\theta_k^0-\theta_k^1)p(x_k), \\ & z^1=\omega^1+\sum_{k=0}^n\theta_k^1 x_k. \end{cases} \tag{25}$$

这里几乎一切都与前面一样，尤其是仍假设只有 n 种风险证券。但假定 $r_1,\dots,r_n$ 之间统计独立，并对期望效用函数 $u=u(x,y)$ 不再作前面那样的假定。在这里我们将假定：

1. u 是光滑函数 (至少有二阶连续偏导数 u_{xx},u_{yy},u_{xy})，并且容许求数学期望与求偏导数的次序交换；
2. u 对两个变量都递增，或者说，对两个变量的偏导数 $u_x,u_y>0$;
3. u 为严格凹函数，即其二阶偏导数所构成的 Hesse 矩阵负定，尤其是 $u_{xx},u_{yy}<0$.

定理 7.7 问题 (25) 的解 $\bar\theta^1$ 及其相应的 $\bar z^0$, $\bar z^1$ 满足下列方程:

$$E\left[u_y(\bar z^0,\bar z^1)(r_j-r_0)\right]=0,\quad j=1,\dots,n. \tag{26}$$

$$r_0=\frac{E\left[u_x(\bar z^0,\bar z^1)\right]}{E\left[u_y(\bar z^0,\bar z^1)\right]}, \tag{27}$$

这里 $r_0,r_1,\dots,r_n$ 分别为各证券的收益率。

证明 证明仍然可用 Lagrange 乘子法。问题 (25) 的 Lagrange 函数为

$$L=E\left[u\left(z^0,\sum_{k=0}^n\theta_k^1x_k\right)\right]+\lambda\left(\omega^0+\sum_{k=0}^n\theta_k^0p(x_k)-z^0-\sum_{k=0}^n\theta_k^1p(x_k)\right).$$

因此，其解 $\bar\theta^1$ 及其相应的 $(\bar z^0,\bar z^1)$ 满足下列条件：

$$\frac{\partial L}{\partial z^0}=0,\quad \frac{\partial L}{\partial\theta_j}=0,\ j=0,1,\dots,n;\quad \frac{\partial L}{\partial\lambda}=0.$$

从而由前两式可得

$$E[u_x(\bar z^0,\bar z^1)]=\lambda, \tag{28}$$

$$E\left[u_y(\bar{z}^0,\bar{z}^1)x_j\right]-\lambda p(x_j)=0,\quad j=0,1,\ldots,n. \tag{29}$$

注意到 $x_j/p(x_j)=r_j$, (29) 可导致

$$E[u_y(\bar{z}^0,\bar{z}^1)x_j/p(x_j)]=E[u_y(\bar{z}^0,\bar{z}^1)(r_j)]=\lambda.$$

取 $j=0$, 再由 (28), 即得 (27) 成立。再取 $j\neq 0$ 时，并与 $j=0$ 的上式相减，即得 (26). □

现在我们来讨论市场均衡。均衡的定义与前面类似。这里我们同样**假定市场均衡存在**。

下一定理说明现在的市场均衡的特点也与前面类似。

定理 7.8 在上述模型中，如果对于定价 $p(x_0)$, $p(x_1)$, …, $p(x_n)$ 来说, $\bar{\theta}^{11}$, $\bar{\theta}^{12}$, …, $\bar{\theta}^{1I}$, 形成市场均衡，那么

i) $\sum_{i=1}^{I}\omega^{0i}=\sum_{i=1}^{I}\bar{z}^{0i}$;

ii) $\bar{\theta}_k^{1i}>0$, $k=1,\ldots,n$, $i=1,2,\ldots,I$.

证明 i) 的证明与定理 7.6 中的情形完全一样。为指出 ii), 由 (26), 可得

$$\operatorname{Cov}\left[u_y^i(\bar{z}^{0i},\bar{z}^{1i}),r_j-r_0\right]=-E\left[u_y(\bar{z}^{i0},\bar{z}^{i1})\right]E[r_j-r_0],\quad j=1,\ldots,n.$$

从而

$$E[r_j-r_0]=-\left(E\left[u_y(\bar{z}^{i0},\bar{z}^{i1})\right]\right)^{-1}\operatorname{Cov}\left[u_y^i(\bar{z}^{i0},\sum_{k=1}^{n}\bar{\theta}_k^{1i}x_k),r_j\right], \tag{30}$$

$$j=1,\ldots,n.$$

这里左端与 i 无关。考虑到 r_j 之间相互独立，则对于 $k\neq j$, $x_k=p(x_k)(r_k)$ 与 r_j 之间也相互独立。因此，由上式可得，如果 $\bar{\theta}_j^{1i}=0$, 则 $E[r_j-r_0]=0$. 另一方面，由中值定理，

$$u_y^i\left(\bar{z}^{i0},\sum_{k=1}^{n}\bar{\theta}_k^{1i}x_k\right)-u_y^i\left(\bar{z}^{i0},\sum_{k\neq j}\bar{\theta}_k^{1i}x_k\right)=u_{yy}^i\left(\bar{z}^{i0},\sum_{k\neq j}\bar{\theta}_k^{1i}x_k+\xi_{x_j}\bar{\theta}_j^{1i}x_j\right)\bar{\theta}_j^{1i}x_j.$$

其中 $\xi_{x_j}\in(0,1)$ (可能依赖于 x_j 的取值)。而 $u_{yy}^i<0$, 故当 $\bar{\theta}_j^{1i}\neq 0$ 时，由上式可得 $E[r_j-r_0]$ 与 $\bar{\theta}_j^{1i}$ 同号。这样就导得所有 $\bar{\theta}_j^{1i}$, $1=1,\ldots,I$ 同号或都等于零。由于对于均衡来说，有 $\sum_{i=1}^{I}\bar{\theta}_j^{1i}=\sum_{i=1}^{I}\theta_j^{0i}>0$, 这只可能所有 $\bar{\theta}_j^{1i}>0$, 即 ii) 成立。 □

现在我们来着手讨论 Ross 的 APT. 讨论的出发点首先仍然是对可数无限种基本证券给出其收益率的模型。与以前一样，假定这无限种证券的收益率分别为 $r_1, r_2, \ldots, r_k, \ldots$, 它们都是随机变量，而 r_0 则是处处取常值的无风险证券的收益率。又假设有 K 种基本因素 $f_1, f_2, \ldots, f_K$ (它们与收益率一样，其当前价格也是 1)，使得下列近似等式成立：

$$r_j - r_0 = \sum_{k=1}^{K} \lambda_{jk}(f_k - r_0) + \varepsilon_j, \qquad j = 1, 2, \ldots, \tag{31}$$

这里假设 f_k 之间相互统计独立， ε_j 是一系列随机“干扰”，它们间相互统计独立，与 f_k 等也独立，以至 $r_1, r_2, \ldots, r_k, \ldots$ 之间相互独立 (这一要求比第五讲中的要高，那里只要求不相关)，并且

$$\mathrm{Var}[\varepsilon_j] \leqslant \sigma^2, \qquad j = 1, 2, \ldots. \tag{32}$$

我们的目的仍然是指出 ε_j 等“很小”。这里比第五讲中更多的假设是：

$$r_j = f_j, \qquad j = 1, \ldots, K; \tag{33}$$

$$\varepsilon_j = \lambda_{j,j+K}(r_{j+K} - r_0), \qquad j = K + 1, \ldots. \tag{34}$$

此外，我们还对每一个投资者对未来消费的 Arrow–Pratt 风险厌恶度量作下列一致有界假设：对未来消费的 Arrow–Pratt 风险厌恶度量有下列一致的界限：

$$-\frac{u^i_{yy}(\bar{z}^{0i}, y)}{E[u^i_y(\bar{z}^{0i}, \bar{z}^{1i})]} \leqslant \gamma^i, \tag{35}$$

其中 y 取任意值， γ^i 为不依赖于 y 的常数。

我们的证明思路是：先假定经济中有 $n > K$ 种风险证券，然后对固定的 n 运用上述的均衡结果，最后再令 $n \to +\infty$, 并指出 $E[\varepsilon_j] \to 0$.

定理 7.9 (APT) 在上述假设下，对于有 n 种证券的市场模型中，如果对于定价 $p(x_0)$, $p(x_1)$, …, $p(x_n)$ 来说, $\bar{\theta}^{11}$, $\bar{\theta}^{12}$, …, $\bar{\theta}^{1I}$, 形成 n 种证券的市场均衡，定义市场组合系数

$$w^M_k(n) = \frac{\sum_{i=1}^{I} \theta^{0i}_k p(x_k)}{\sum_{k=1}^{n} \sum_{i=1}^{I} \theta^{0i}_k p(x_k)}, \quad k = 1, \ldots, n,$$

那么，当

$$\lim_{n \to \infty} w^M_k(n) = 0, \qquad k = K + 1, \ldots, n;$$

时，有

$$\lim_{n\to\infty} E\left[r_j - r_0 - \sum_{k=1}^{K} \lambda_{jk}(f_k - r_0)\right] = \lim_{n\to\infty} E[\varepsilon_j] = 0, \quad j = 1, 2, \ldots, n. \tag{36}$$

证明 与上一定理的证明一样，我们由 (30), 对 u_y 用中值定理以及 (35) 和 $\bar{\theta}_j^{1i} > 0$, 可导得

$$E[r_j - r_0] \leqslant \gamma^i \bar{\theta}_j^{1i} p(x_j) \mathrm{Cov}[r_j, r_j], \quad j = 1, \ldots, n.$$

从而由 (34), 当 $j > K$ 时，由 (32) 可得

$$E[\varepsilon_j] = E[\lambda_{j,j+K}(r_{j+K} - r_0)] \leqslant \gamma^i \bar{\theta}_{j+K}^{1i} p(x_{j+K}) \frac{\mathrm{Var}[\varepsilon_j]}{|\lambda_{j,j+K}|} \leqslant \frac{\sigma^2 \gamma^i}{|\lambda_{j,j+K}|} \bar{\theta}_{j+K}^{1i} p(x_{j+K}).$$

尤其是由于在市场均衡时，可导得

$$E[\varepsilon_j] \leqslant \frac{\alpha}{|\lambda_{j,j+K}|} \frac{(\sum_{i=1}^{I} \bar{\theta}_{j+K}^{1i}) p(x_{j+K})}{\sum_{j=1}(\sum_{i=1}^{I} \bar{\theta}_j^{1i}) p(x_j)} = \frac{\alpha}{|\lambda_{j,j+K}|} w_{j+K}^M(n),$$

其中 α 是常数。由此即可导得定理的结论。 □

这样，我们就从均衡定价的角度，给出了 Ross 的 APT 的新论证。与前面第五讲中所得到的结果相比较，前面得到的是

$$\lim_{n\to\infty} \sum_{j=1}^{n} (E[\varepsilon_j])^2 < +\infty;$$

而现在得到的是

$$\lim_{n\to\infty} E[\varepsilon_j] = 0, \quad j = 1, 2, \ldots, n.$$

显然现在的结论更强。但是为得到这一更强的结论是有代价的。均衡定价除了其框架上的许多一般假设外，尤其是还假设“投资者期望均匀性”；而第五讲中的讨论没有这一假设，它可以根据某一投资者的个人信念 (对事件发生的概率的估计) 来展开。定理 7.9 中的 $\lim_{n\to\infty} w_k^M(n) = 0$ 也是一项重要假设。它意味着除前 K 个因子外，在市场组合中的每一证券所占的比例都是越来越小的。因此，有 $E[\varepsilon_j]$ 一致地小也是不令人惊奇的。此外，前一式作为一个单独的条件比较容易检验，而后一

式代表许多极限式，不太容易检验。至于 APT 的实际检验研究，我们在前面已经提到，这里不再重复。

从以上的讨论中，我们还可看到 CAPM 和 APT 的不同是因为它们的假设不同，尤其是对期望效用函数的假设不同。CAPM 是一个单因子模型，并且这一因子还被要求为有效的均衡市场组合的收益率；这是它被批评为无法检验的根本弱点。而 APT 是一个多因子模型，对因子本身的有效性没有要求。这正是 Ross 认为 APT 是比 CAPM 更好的替代物的论据。从实际应用来说，APT 看来确实更好。但是 APT 的多因子模型形式是作为假设引进的，这样的先验假设理论上说不出更多的理由。因此，对于金融经济学的理论研究 (例如关于 Modigliani–Miller 理论的深入研究) 来说，CAPM 的框架还是比 APT 框架更常用。

说明 正如我们在“代引言”中所说，Arrow–Debreu 一般经济均衡存在定理应该是现代金融经济学的基础。无论是他们的经典论文 [1]，还是 Debreu 的经典著作 [2]，至今仍然值得一读。有些简单的讨论，可参看 Varian [17]，其中尤其包括 Brouwer 不动点定理与 Debreu–Gale–Nikaïdo 定理等价的证明。纯交换经济模型虽然从经济内容看来比较简单，但是已经包含一般情形中的几乎所有要素。对于金融市场均衡的讨论，更是从纯交换经济模型出发更为方便。我们的叙述很大程度上取材自 Magill–Quinzii [11]。Radner [14] 原来提出的模型包含多种商品，而不是只有“钱”(单商品)。于是模型中还可包括期货合约等等。这种多商品的模型虽然更为复杂，但在数学上只是简单的推广。这里最重要的还是市场的不完全性使问题起质的变化。目前仅对有限状态的不完全市场的一般均衡理论讨论得比较清楚。无限状态的不完全市场的一般均衡理论有许多难以克服的数学困难。

关于金融市场一般均衡的讨论在 Huang–Linzenberger [8] 中见其第 5 章；在 Ingersoll [9] 中见其第 8 章；在 Jarrow [10] 中则有第 6、15 和 18 章。看来最后这本教材对此最重视。不过 Duffie 的 [3], [4] 倒是一开始就强调一般经济均衡，其中 [3] 甚至把经典的 Arrow–Debreu 定理都概要地进行了全面介绍。我们的最后两小节的叙述比较接近 Jarrow 的书。最后一节有关 Ross 的 APT 的叙述实际上与 Ross 的经典论文 [15] 基本一致。

Markowitz [12] 中用相当大的篇幅来讨论期望效用函数，目的是要为他的均值－方差分析建立“经济学基础”。但是其最终结果只是说他的理论可以用 7.4 中那样的效用函数来阐述。实际上，正如我们在上一讲中已经提到的，从根本上来说，Markowitz 均值－方差准则并不是期望效用函数的特殊情形。Sharpe 的有关 CAPM 的经典论文 [15] 中虽然不断提到市场均衡，但是 CAPM 的一般均衡模型其实是后来提出的。最早比较明确提出的是 Duffie [3]。后来的研究有例如，Nielsen [13] Magill–Quinzii [11] 则更是把它与不完全市场一般均衡理论联系起来。

思考题和习题

1. 什么是纯交换经济？它的数学模型是怎样的？

2. 什么是纯交换经济的一般经济均衡存在定理和福利经济学第一基本定理？其中对消费者的效用函数有什么假设？

3. 怎样通过资产定价基本定理把完全金融市场的均衡存在定理归结为纯交换经济的均衡存在定理？比较这两条在经济内容上非常不同的定理，你对完全金融市场的均衡有什么看法？

4. 什么是不完全金融市场？为什么它不能转化为纯交换经济那样的形式？对于不完全金融市场的一般均衡有什么引入注目的结论？

5. 怎样利用均值－方差型的效用函数来给 CAPM 提出一般经济均衡框架？它在理论上导出哪些结论？其模型有什么不足？

6. 怎样从一般经济均衡框架来导出 Ross 的 APT 理论？它有哪些基本假设？其结论与第五讲中介绍的有何不同？

参考文献

[**1**] Arrow, K., and G. Debreu, 1954, Existence of an equilibrium for a competitive economy, *Econometrica*, **22:** 265–290.

[**2**] Debreu, G., 1959, Theory of Value, An Axiomatic Analysis of Economic Equilibrium, Yale University Press, New Haven and London.

[**3**] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[**4**] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N. J..

[**5**] Duffie, D., and W. Schafer, 1985, Equilibrium in incomplete markets I: Basic model of generic existence, *Journal of Mathematical Economics*, **14**: 285–300.

[**6**] Duffie, D., and W. Schafer, 1986, Equilibrium in incomplete markets II: Generic existence in stochastic economies, *Journal of Mathematical Economics*, **15**: 199–216.

[**7**] Hart, O. D., On the optimality of equilibrium when the market structure is incomplete, *Journal of Economic Theory*, **11**: 418–443.

[**8**] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[**9**] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[**10**] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[**11**] Magill, M. & M. Quinzii, 1997, Theory of Incomplete Markets, Vol. 1, MIT Press, Cambridge, Massachusetts.

[**12**] Markowitz, H., 1959, 1991 Second ed., Portfolio Selection: Efficient Diversification of Investment, Basil Blackwell, Cambridge.

[**13**] Nielsen, L. T., 1990, Existence of equilibrium in CAPM, *Journal of Economic Theory*, **52**: 223–231.

[**14**] Radner, R., 1972, Existence of equilibrium of plans, prices and prices and price expections in a sequence of markets, *Econometrica*, **40**: 289–304.

[**15**] Ross, S. A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, **13:** 341–360.

[**16**] Sharpe, W., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, **19:** 425–442.

[**17**] Varian, H., 1992, Microeconomic Analysis, 3rd ed., W. W. Norton, New York.

八、Black–Scholes 期权定价理论

本讲要求：熟悉 Black–Scholes 期权定价公式及其计算。了解 Black–Scholes 公式与资产定价基本定理之间的关系。了解 Black–Scholes 公式的 Cox–Ross–Robinstein (二叉树方法) 推导思路。对一般的多期模型了解资产定价基本定理的新形式，特别是鞅的概念。

数学预备知识：除一般高等数学知识外，需要概率论中的中心极限定理，它的证明涉及随机变量的特征函数等概念。对于数学基础不足的读者，可以通过理解最后结果来接受它。本讲中还会涉及随机序列、鞅等概念，但有关知识也将在本讲中叙述。

内容：

- 8.1 Black–Scholes 欧式买入期权定价公式。
- 8.2 Black–Scholes 公式的前驱。
- 8.3 Black–Scholes 公式的 Cox–Ross–Robinstein (二叉树方法) 推导。
- 8.4 一般的有限状态多期模型。
- 8.5 资产定价基本定理的新形式。鞅的概念。
- 8.6 更一般的多期模型。与线性定价法则的联系。

这一讲将讨论 Black–Scholes 期权定价理论。其基本思想在第一讲中已经介绍过。但是在那里我们主要是为了说明无套利假设、未定权益复制、风险对冲等基本概念，并没有导出最终结果。这里我们首先要继续第一讲的讨论，最终导出著名的 Black–Scholes 欧式期权定价公式。为导出这一公式，我们仍然只需要“套利定价论”，而不需要“均衡定价论”。

8.1　我们先来看看什么是 Black–Scholes 期权公式。经典的 Black–Scholes 期权定价公式是对于欧式股票期权给出的。其公式为

$$C(S,T)=SN(d_1)-Ke^{-rT}N(d_2),$$

其中 T 是到期时间，S 是当前股价，$C(S,T)$ 是作为当前股价和到期时间的函数的欧式买入期权的价格；

$$d_1=\frac{1}{\sigma\sqrt{T}}\left[\log\frac{S}{K}+\left(r+\frac{\sigma^2}{2}\right)T\right],\qquad d_2=d_1-\sigma\sqrt{T},$$

K 是期权的执行价格，r 是无风险证券的 (瞬时) 收益率，σ 称为股价的**波动率** (volatility，(连续时间) 收益率的标准差与时间间隔的开方之比，这是一个需要测算的参数)；N 称为累积正态分布函数，定义为

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-\frac{y^2}{2}} dy.$$

这个公式当 $T = 0$ 时，有 $C(S, 0) = (S - K)_+$, 其中 $(S - K)_+$ 表示 $S - K$ 的正部，即当 $S - K \geqslant 0$ 时，它等于 $S - K$, 但当 $S - K < 0$ 时，它为零。这就是买入期权在到期时的价值，它取决于股价与执行价格之差。而当 $T > 0$ 时，$C(S, T)$ 作为 S 的函数，其图象逐渐上移，但原点不动。当 T 趋向 $+\infty$ 时，图象趋向于 $C = S$（如图 1)。

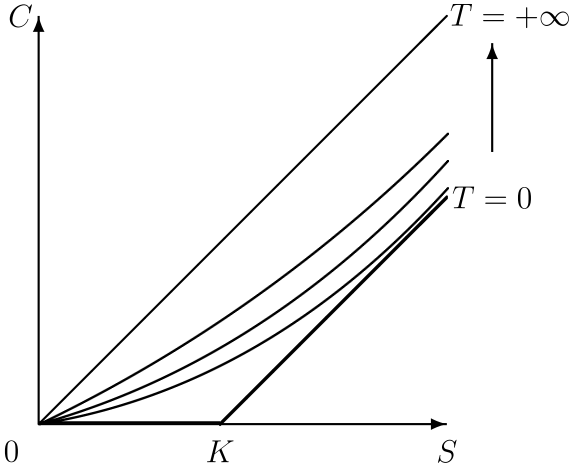


图 1　期权价格曲线随到期时间 T 的变化

Black–Scholes 公式的方便之处在于除股价的波动率 σ 外，其他参数都是直接在市场上可以找到的。虽说在严格的意义下，r 是无风险证券的瞬时收益率，但是在实际计算中，它直接可以用短期利率代入。例如，如果 $S = 18$, $K = 15$, $r = 10\%$, $T = 0.25$, $\sigma = 15\%$, 这里价格以美元计，时间以年计，从而涉及的两个比率都指的是年率，那么 (以下的等号实际上都是近似等号)

$$Ke^{-rT} = 15e^{-0.1(0.25)} = 14.6296,$$

$$d_1 = \frac{\log(18/15) + [0.1 + (0.15)^2(0.5)]0.25}{0.15\sqrt{0.25}} = \frac{0.21013}{0.075} = 2.8017,$$

$$d_2 = d_1 - 0.15\sqrt{0.25} = 2.7267.$$

把这些值代入公式，得到

$$C = 18N(2.8017) - 15\, e^{-0.1(0.25)} N(2.7267).$$

利用累积正态分布函数在点 2.8017 和 2.7267 处的近似值，买入期权的价格是 3.3749，即

$$C = 18 \cdot (0.997) - 14.6296 \cdot (0.996) = 3.3749.$$

更精确的计算可得 $C = 3.3714$.

8.2 Black–Scholes 公式的出现不是偶然的。历史上曾经有过许多类似的公式。其中最早可追溯到 1900 年法国数学家 Bachelier (1870–1946) 的博士论文。这些公式中有的在形式上已经非常接近，但是都因在机理上没有击中要害，没有达到 Black–Scholes 的高度。我们下面列举一些历史上曾经有过的一些公式，其中所有记号的意义都与前面一样，新增加或变化的记号则另行说明。

1. Bachelier (1900)：

$$C(S,T) = SN\left(\frac{S-K}{\sigma\sqrt{T}}\right) - KN\left(\frac{S-K}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}n\left(\frac{K-S}{\sigma\sqrt{T}}\right)$$

   n 是正态分布的密度函数。

2. Sprenkle (1961)：

$$C(S,T) = e^{\rho T} SN(d_1) - (1-A)KN(d_2)$$

   其中

$$d_1 = \frac{1}{\sigma\sqrt{T}}\left[\log\frac{S}{K} + \left(\rho + \frac{1}{2}\sigma^2\right)T\right], \qquad d_2 = d_1 - \sigma\sqrt{T},$$

   ρ 是股票价格的平均增长率，A 对应风险厌恶程度。

3. Boness (1964)：

$$C(S,T) = SN(d_1) - Ke^{-\rho T}N(d_2),$$

   其中

$$d_1 = \frac{1}{\sigma\sqrt{T}}\left[\log\frac{S}{K} + \left(\rho + \frac{1}{2}\sigma^2\right)T\right], \qquad d_2 = d_1 - \sigma\sqrt{T}.$$

4. Samuelson (1965)：

$$C(S,T) = Se^{-(\rho-\alpha)T}N(d_1) - Ke^{-\alpha T}N(d_2)$$

其中

$$d_1 = \frac{1}{\sigma\sqrt{T}}\left[\log\frac{S}{K} + \left(\rho + \frac{1}{2}\sigma^2\right)T\right], \qquad d_2 = d_1 - \sigma\sqrt{T}.$$

α 是期权价格的平均增长率。

我们可以看到，所有这些公式都与后来的 Black–Scholes 公式有许多相似的地方。但是所有这些公式中都有一个或多个参数，它们依赖于投资者对风险或对股票收益率的偏好等。相反， Black–Scholes 公式没有这样的参数。然而， Black–Scholes 公式中有一个前人没有考虑的参数，那就是无风险证券的收益率 r. 这是因为前人都没有考虑资金的“时间价值”。甚至有人已经考虑到风险完全对冲的组合，但由于没有注意到无风险证券组合的收益率应该等于无风险证券的收益率，而功亏一篑。

8.3 下面我们来正式推导 Black–Scholes 公式。这一推导方法是 Cox–Ross–Rubinstein [5] 于 1979 年提出的[1]。在这一推导过程中，我们只用到资产定价基本定理，也就是说只用到无套利假设，其他都是一些具体的数学运算。至于 Black–Scholes 原来的推导方法我们将在以后介绍。

我们先来回顾第一讲的讨论。在那里我们假定股票的当前价格为 S_0, 股票的未来价格有两种可能性： aS_0 和 bS_0; 期权的执行价格为 K. 那么期权在“未来”可能有的两种价格为 $C_a = \max(aS_0 - K, 0)$ 和 $C_b = \max(bS_0 - K, 0)$. 在无套利假设下，有一种可能性估计，使得“未来”的股价的平均值恰好就等于“当前”的股价。这是因为 a 和 b 中必然有一个大于 1, 另一个小于 1, 从而存在 0 与 1 之间的数 p (概率) 使得

$$ap + (1-p)b = 1, \qquad \text{即} \qquad p = \frac{1-b}{a-b}.$$

而“当前”的期权价格应该就是在这种可能性估计下的“未来”的期权价格的平均值，即

$$C_0 = pC_a + (1-p)C_b = \frac{C_a - C_b + aC_b - bC_a}{a-b}.$$

---

[1] 1979 年 Rendleman–Bartter 发表的 [19] 也有类似的结果。该文中说，他们是独立于 [5] 发现的，并且指出 Sharpe 当时也已有类似的思想。

把这个模型略为复杂化，我们假定无风险 (净) 利率[2]为 R. 那么当前的一元钱到了未来变为 $(1+R)$ 元。这时无套利假设 (资产定价基本定理) 将导至：

$$pa + (1-p)b = 1 + R. \tag{1}$$

由此导得 $p = \frac{1+R-b}{a-b}$, 而公式将是

$$C_0 = \frac{1}{1+R}\left(pC_a + (1-p)C_b\right) = \frac{(1+R)(C_a - C_b) + aC_b - bC_a}{(1+R)(a-b)}.$$

现在我们把模型再进一步复杂化，假定所考虑时间有 $0, 1, \ldots, N$ 个时刻，每个时刻的短期银行利率仍为 R. 那么当前的一元钱到 k 时刻后就变为 $(1+R)^k$ 元钱，$k = 1, 2, \ldots, N$. 同时又假定每个时刻的股价变化仍然只有两种可能，即隔时刻间的股价比或是 a, 或是 b, 其中 $a < 1 + R < b$. 但是 N 时刻后的股价变化状态有 2^N 种。这样的变化可以如图 2 中那样来表示。这是一个在每一时刻的每一状态都有两个分叉的图。因而这种方法通常称为**二叉树方法**。

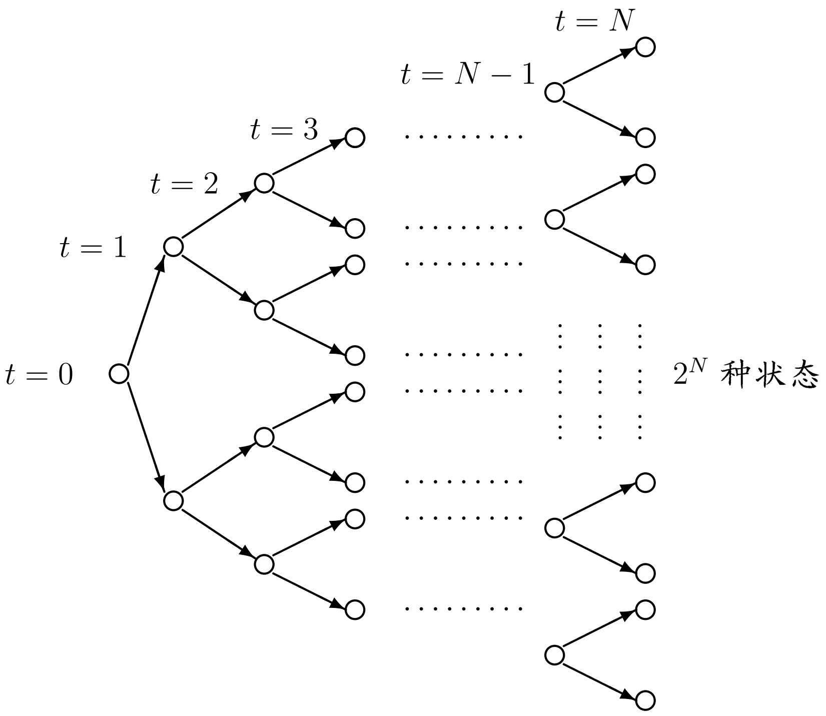


图 2 二叉树的状态变化图

---

[2] 本讲中的无风险利率都是指“净利率”，与我们以前为讨论方便所引入的“总利率”相差 1。

这 2^N 种状态的股价与当前的股价的比值有些是一样的，它们分别是

$$\begin{aligned}&C_N^0 \text{ 个 } a^N,\\ &C_N^1 \text{ 个 } a^{N-1}b,\\ &\cdots\cdots\cdots,\\ &C_N^k \text{ 个 } a^{N-k}b^k,\\ &\cdots\cdots\cdots,\\ &C_N^N \text{ 个 } b^N.\end{aligned} \tag{2}$$

因此，如果把比值一样的状态看成一样的状态，它的状态图将如图 3 (由于它与二项式定理的展开是完全一样的，不妨称它为“二项式状态变化图”)：

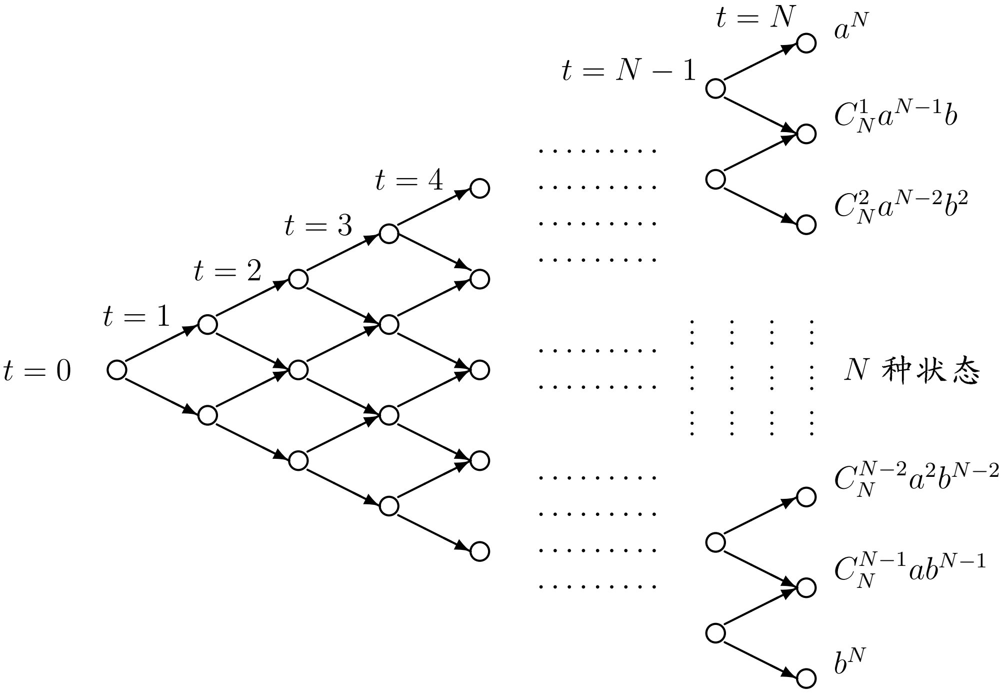


图 3　二项式状态变化图

(欧式) 买入期权 (call option) 是在 N 时刻后，以执行价格 K 买入一份股票的权利。这种衍生证券我们目前不知道它当前应该值多少钱，但是 N 时刻后，股价 S_N 确定后，它的价格也就确定了，即 (S_N-K) 的正部 $(S_N-K)_+$, 它在 $S_N \geqslant K$ 时等于 (S_N-K), 而在 $S_N<K$ 时为零，因为这时期权持有者自然不会去执行他的权利。类似的是 (欧式) 卖出

期权 (put option), 它是在 N 时刻后以某种固定价格 K 卖出一份股票的权利。其 N 时刻后的价格是 $(K-S_N)_+$. 它们都是在 N 时刻后有 2^N 种可能的未定权益 (衍生证券，数学上是 2^N 维向量 $(d_k)_{k=1,\ldots,2^N}$)。它们的定价问题就是要问它当前值多少。这个问题我们已经知道可用与无套利假设等价的资产定价基本定理来回答。这里我们就可利用上述的等价概率测度。这一概率测度使得股价收益率变化的平均值始终与无风险收益率 (银行利率) 一样，从而由它来确定衍生证券价格的平均值就是它当前的 (无套利) 价格。因此，衍生证券当前的价格 $\bar{d}$ 应该等于

$$\bar{d} = \frac{1}{(1+R)^N} E^*\left[(d_k)\right]$$

$$= \frac{1}{(1+R)^N}\left((1-p)^N d_1 + \sum_{j=1}^{C_N^1}(1-p)^{N-1}p d_{1+j} + \sum_{j=1}^{C_N^2}(1-p)^{N-2}p^2 d_{1+C_N^1+j} \right.$$

$$\left. + \cdots + \sum_{j=1}^{C_N^{N-1}}(1-p)p^{N-1}d_{1+C_N^1+\cdots+C_N^{N-2}+j} + p^N d_{2^N}\right), \qquad (3)$$

其中 E^* 表示对随机变量 (d_k) 按上述等价概率测度求数学期望，它就等于右端的求和值。前面的因子 $1/(1+R)^N$ 则是把它折现为按当前的“不变价格”来计算。

事实上，在上述模型下，欧式买入期权的当前价格 $C_0^{(N)}$ 就应该是

$$C_0^{(N)} = \frac{1}{(1+R)^N}\sum_{k=0}^{N} C_N^k p^{N-k}(1-p)^k \left(Sa^{N-k}b^k - K\right)_+, \qquad (4)$$

这里 S 是当前的股价 (我们省略了下标 0)，而 N 时刻后的股价是随机变量 (2^N 维向量) S_N, 它的 2^N 个分量由 (2) 式中的因子与当前的股价 S 的乘积来确定。同样，欧式卖出期权的当前价格 $P_0^{(N)}$ 应该是

$$P_0^{(N)} = \frac{1}{(1+R)^N}\sum_{k=0}^{N} C_N^k p^{N-k}(1-p)^k \left(K - Sa^{N-k}b^k\right)_+. \qquad (5)$$

由 (4), (5), (1) 以及二项式定理，立即可得

$$C_0^{(N)} - P_0^{(N)} = \frac{1}{(1+R)^N}\sum_{k=0}^{N} C_N^k p^{N-k}(1-p)^k \left(Sa^{N-k}b^k - K\right) = S - \frac{K}{(1+R)^N}. \qquad (6)$$

它称为**买入－卖出期权平价关系** (call-put parity)。这个等式也可由无套利假设直接导出；即当这个关系式不满足时一定存在套利机会。下面我们将通过 $N \to \infty$ 来导出 Black–Scholes 公式。

为此我们假设时间的总长度是 T, 它被 N 等分。利率 $R = rT/N$, 其中 r 是“瞬时利率”，即

$$e^{rT} = \lim_{N\to\infty}(1+R)^N = \lim_{N\to\infty}\left(1+\frac{rT}{N}\right)^N.$$

又设

$$\log\frac{a}{1+R} = -\frac{\sigma_T}{\sqrt{N}}, \qquad \log\frac{b}{1+R} = \frac{\sigma_T}{\sqrt{N}}. \tag{7}$$

这样的假设对 a, b 有特殊要求。但是这里只是为了简化计算。事实上，不作这样的特殊假设并不影响下面的计算结果。可以证明， σ_T^2 是随机变量 $\log(S_N)$ (或者说，取值 a 和 b 的二项分布随机变量，它与 $\log(S_N)$ 差一个常数) 的方差当 $N \to \infty$ 时的极限。当 $N \to \infty$ 时，我们需要二项分布趋向于正态分布的特殊中心极限定理；即

引理 8.1 设 (Y_N) 是如下形式的随机变量列：

$$Y_N = X_1^N + X_2^N + \cdots + X_N^N,$$

其中对于每个 N, 随机变量 X_i^N, $i = 1, \ldots, N$, 相互独立，有同样的分布，且都只取两个值 $\{-\sigma_T/\sqrt{N}, \sigma_T/\sqrt{N}\}$, 以及有满足 $\lim_{N\to\infty}(N\mu_N) = \mu$ 的数学期望值 μ_N. 那么序列 (Y_N) 按分布趋向于均值为 μ, 标准差为 σ_T 的正态随机变量。

为证明这一定理，只需考虑 Y_N 的特征函数 ϕ_{Y_N}:

$$\phi_{Y_N}(u) = E\left[e^{iuY_N}\right] = \prod_{j=1}^{N} E\left[e^{iuX_j}\right]$$

$$= \left(E\left[e^{iuX_1^N}\right]\right)^N = \left(1 + iu\mu_N - \frac{\sigma_T^2 u^2}{2N} + o(1/N)\right)^N.$$

因此， $\lim_{N\to\infty}\phi_{Y_N}(u) = \exp(iu\mu - (\sigma_T^2 u^2/2))$; 从而引理成立。

现在我们取

$$X_j^N = \log\left(\frac{T_j}{1+R}\right), \qquad Y_N = \sum_{j=1}^{N}\log\left(\frac{T_j}{1+R}\right),$$

其中 T_j 是取 a 或 b 的二项分布随机变量。考虑到上面的对 a, b 的假设，X_j^N 符合定理假设。把 (5) 式中定义的平均值理解为以 E^* 表示的数学期望，则它可以重记为

$$\begin{aligned}
P_0^{(N)} &= \left(1+\frac{rT}{N}\right)^{-N} E^*\left[\left(K - S\prod_{j=1}^{N} T_j\right)_+\right] \\
&= E^*\left[\left(\left(1+\frac{rT}{N}\right)^{-N} K - Se^{Y_N}\right)_+\right].
\end{aligned}$$

另一方面，由 (7), 我们又有

$$E^*[X_j^N] = (1-2p)\frac{\sigma_T}{\sqrt{N}} = \frac{2 - e^{\sigma_T/\sqrt{N}} - e^{-\sigma_T/\sqrt{N}}}{e^{\sigma_T/\sqrt{N}} - e^{-\sigma_T/\sqrt{N}}}\frac{\sigma_T}{\sqrt{N}},$$

从而

$$\lim_{N\to\infty} NE^*[X_j^N] = -\frac{\sigma_T^2}{2}.$$

这样，令 $\psi(y) = (Ke^{-rT} - Se^y)_+$, 可得

$$\begin{aligned}
&\left|P_0^{(N)} - E^*[\psi(Y_N)]\right| \\
&= \left|E^*\left[\left(\left(1+\frac{rT}{N}\right)^{-N} K - Se^{Y_N}\right)_+ - \left(Ke^{-rT} - Se^{Y_N}\right)_+\right]\right| \\
&\leqslant K\left|\left(1+\frac{rT}{N}\right)^{-N} - e^{-rT}\right| \to 0.
\end{aligned}$$

利用上述引理 8.1，并注意到 ψ 是有界的连续函数，就有

$$\lim_{N\to\infty} P_0^{(N)} = \lim_{N\to\infty} E^*\left[\psi(Y_N)\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(Ke^{-rT} - Se^{-\frac{\sigma_T^2}{2}+\sigma_T y}\right)_+ e^{-\frac{y^2}{2}}dy.$$

再令

$$N(d) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d} e^{-\frac{x^2}{2}}dx,$$

$$d_1 = \frac{\log\frac{S}{K} + rT + \frac{\sigma_T^2}{2}}{\sigma_T}, \qquad d_2 = d_1 - \sigma_T,$$

最后得到

$$\lim_{N\to\infty} P_0^{(N)} = Ke^{-rT}N(-d_2) - SN(-d_1),$$

以及

$$\lim_{N\to\infty} C_0^{(N)} = \lim_{N\to\infty}\left(P_0^{(N)} + S - K\left(1+\frac{rT}{N}\right)^{-N}\right) = SN(d_1) - Ke^{-rT}N(d_2).$$

即以时刻 T 为到期日、当前股价为 S 的两种欧式期权的当前价格 $c(S,0)$[3] 和 $p(S,0)$ 分别为

$$c(S,0) = SN(d_1) - Ke^{-rT}N(d_2), \quad p(S,0) = Ke^{-rT}N(-d_2) - SN(-d_1).$$

这就是 **Black–Scholes 期权定价公式**。

更一般的 Black–Scholes 公式为

$$c(S_t,t) = S_tN(d_1^t) - Ke^{-r(T-t)}N(d_2^t),$$

$$p(S_t,t) = Ke^{-r(T-t)}N(-d_2^t) - S_tN(-d_1^t),$$

其中 $t \in [0,T]$, 而

$$d_1^t = \frac{\log\frac{S_t}{K} + (r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \qquad d_2^t = d_1^t - \sigma\sqrt{T-t}.$$

它们表示在任意时刻的期权价格与股价 (它是随机变量！) 之间的关系。其证明可用完全一样的思路。

这样，我们就导得了著名的 Black–Scholes 期权定价公式。虽然其中用到了一些较细腻的数学技巧。但是我们可以看到，其基本思想与我们前面讨论的“当前－未来”的二期模型中的用到的没有本质区别。最关键的结果还是与无套利假设等价的资产定价基本定理。上面的证明不但提供了 Black–Scholes 公式的一种证明方法，同时，它也提供了一种计算方法。事实上，利用二叉树方法，确实也能给出期权定价的具体计算。

8.4　以下我们要把上述讨论推广到一般的有限状态多期模型情形。这是 Black–Scholes 期权定价理论的二叉树方法的一般化。它们更多地来自我们曾提到过的 Harrison–Kreps [9] 的工作。多期与二期的区别仅在于讨论中涉及的空间维数有所增加。最终得到的结论将是完全类似的。

与二期模型相比的本质不同在于它常常要从不确定的“明天”出发来考虑更不确定的“后天”等等。于是其中涉及的不确定性将随着时间

---

[3] 这里的期权价格 $c(S,t)$ 相当于第一节中的 $C(S,T-t)$. 同时，$\sigma_T = \sigma\sqrt{T}$.

的向前推移越来越复杂。就象我们在上面所看到的那样，经过 N 个时刻以后，可能的状态会从 1 种， 2 种，一直展开到 2^N 种。这里关键的概念是所谓“事件树”或“信息流”。

假设现在我们考虑的问题有 $N+1$ 个时刻。从当前的确定状态出发，下一时刻可能发生若干种状态，从下一时刻可能的每一种状态出发，再下一时刻又会发生若干种状态。如此等等，一直到 N 个时刻后，所有可能发生的状态形成一个有限集 (我们这里只考虑有限) Ω. 整个这样的过程形成一棵“事件树”(又称“信息流”)，其中每一个节点都代表某时刻的某个状态；而从每个节点出发，它又都能生成该“事件树”的一部分，这部分的终点仍都落入 Ω 内。同时，从不同的节点出发所生成的“事件树”部分也都是不同的，并且互不交叉。前面的图 2 是每个节点只生成两种新状态的“事件树”的特殊情形。

整个这样的“事件树”代表着一股信息流。记 $\Omega_0, \Omega_1, \ldots, \Omega_N = \Omega$ 为代表每时刻的各个可能的状态节点集合，其中 Ω_0 代表当前的状态集，应该只有一个元素。记 Ω 的所有子集所形成的集合为 $\mathcal{F}$, 以及一般记 Ω_n 的所有子集所形成的集合为 $\mathcal{F}_n$, 并把 Ω_n 中的每一节点与它在 Ω_{n+k} $(k \leqslant N-n)$ 中可达到的节点全体看作一样，那么就有

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_{N-1} \subset \mathcal{F}_N = \mathcal{F}.$$

这一事实在概率论中有个专门名词称为 **σ-域流** (filtration). 其定义比这里要一般得多。“事件树”所生成的 σ-域流不但是一般的 σ-域流的最简单的情形，也为一般的 σ-域流提供了一个直观形象。

我们不妨把 Ω_n 与它所包含的元素个数混为一谈。那么在上述模型中，一个**随机变量**就是 $\mathbb{R}^\Omega = \mathbb{R}^{\Omega_N}$ 中的向量，即它对每个最终状态都有一个数值。同样，对于任何 $n \leqslant N$, $\mathbb{R}^{\Omega_n}$ 中的元素也可看作一个随机变量，只是它被理解为对于每个从 Ω_n 中的状态出发能在 Ω 中达到的所有状态上取同样的值。用概率论的术语来说，这样的随机变量称为是 **$\mathcal{F}_n$-可测的**，其含义是这种随机变量在第 n 时刻时已包含其全部信息。尤其是 $\mathcal{F}_0$-可测的随机变量也就是当前的唯一状态下的确定的量。

至此，我们还没有引进概率。所谓概率测度 p 是 $\mathbb{R}^\Omega_{++}$ (所有分量严格大于零的 $\mathbb{R}^\Omega$ 中的向量) 中的元素，且它满足 $\sum_{\omega\in\Omega} p(\omega) = 1$. 每个这样的概率测度都被称为 Ω 上的等价概率测度，因为它们都对可能发生的状态给予非零概率。对于随机变量 $z \in \mathbb{R}^\Omega$,

$$E[z] = p \cdot z = \sum_{\omega\in\Omega} p(\omega) z(\omega)$$

称为 z 的**数学期望**。它是随机变量各种可能取的值按其可能性大小来取的平均值。数学期望是一个确定的数值。或者说是“当前的数值”。

有时我们也要考虑一个随机变量在第 n 时刻的平均值，由于第 n 时刻的状态在当前看来也是不确定的，从而这个平均值将是一个 $\mathcal{F}_n$-可测的随机变量。这种平均值称为**条件数学期望**。一个 $\mathcal{F}_n$-可测的随机变量就是一个 $\mathbb{R}^{\Omega_n}$ 中的一个向量，其中每个 $\omega_n \in \Omega_n$ 都可理解为 Ω 中的一个子集，即从 ω_n 出发的在 Ω 中可达的状态节点全体。一个随机变量也可在这个状态节点集上求平均值。于是随机变量 $z \in \mathbb{R}^{\Omega}$ 关于 $\mathcal{F}_n$ 的条件数学期望定义为

$$E[z|\mathcal{F}_n] = (z(\omega_n))_{\omega_n \in \Omega_n} \in \mathbb{R}^{\Omega_n},$$

$$z(\omega_n) = \frac{\sum_{\omega \in \omega_n} p(\omega) z(\omega)}{\sum_{\omega \in \omega_n} p(\omega)}.$$

当然，数学期望和条件数学期望的定义都取决于事先给出的概率测度。

若干个随机变量按次序放在一起考虑称为**随机序列**或**随机过程**。在我们考虑的问题中，随机过程总是与各时刻联系在一起的，因此，我们考虑的随机过程总是有下列形式：

$$x = (x_n)_{n=0,1,\ldots,N.}$$

并且我们以后考虑的每个 x_n 都将是多维的随机变量，(x_n) 是多维随机过程。有一种随机过程 (S_n), 它在第 n 时刻时的值对第 n 时刻来说已完全确定，即 $S_n \in \mathbb{R}^{\Omega_n}$ 是 $\mathcal{F}_n$-可测的，那么它称为**适应过程**；以下讨论的问题中，证券的价格过程就是适应过程，因为它在第 n 时刻的每一状态下的值都是确定的。还有一种随机过程 (ϕ_n), 除了当前外，它在第 n 时刻时的值对第 $n-1$ 时刻来说已完全确定，即 $\phi_n \in \mathbb{R}^{\Omega_{n-1}}$ 是 $\mathcal{F}_{n-1}$-可测的，那么它称为**可料过程**。以下讨论的问题中，投资者每时刻进行的证券组合交易就是可料过程，因为第 n 时刻起作用的证券组合交易是在第 $n-1$ 时刻时作的。

有了这些数学准备以后，我们就可来讨论我们的证券交易模型。假定模型中有 $K+1$ 种证券，其中第 0 种证券是无风险证券，即其每时刻的价格是一个确定过程 $S_0^0 = 1$, $S_1^0 = (1+R)$, $\ldots$, $S_N^0 = (1+R)^N$, 其中 R 是短期利率。其他 K 种证券都是风险证券，它们每时刻的价格对于每时刻不同的状态都是确定的，即它们都是所谓适应过程。记**价格过程**为

$$S_n = (S_n^0, S_n^1, \ldots, S_n^K), \quad n = 0, 1, \ldots, N.$$

投资者每时刻进行的证券交易是一个向量 $\phi_n = (\phi_n^0, \phi_n^1, \ldots, \phi_n^K)$, 其中每个分量表示对该种证券的交易量。$\phi = (\phi_n)$ 称为投资者的**策略过程**，作为随机过程它是可料过程，因为对第 n 时刻起作用的策略是在第 $n-1$ 时刻作出的。

投资者在第 n 时刻时手中掌握的证券是 ϕ_n (它在不同的状态下是不同的)。由于当时的证券价格为 S_n, 故这一证券组合的价值为

$$V_n(\phi) = \phi_n \cdot S_n = \sum_{i=0}^{K} \phi_n^i S_n^i.$$

如果我们都以当前的“不变价格”来折现，令 $\tilde{S}_n = S_n/(1+R)^n$ (我们以后称它为**折现价格**), 那么这一证券组合的“折现价值”应该是

$$\tilde{V}_n(\phi) = (\phi_n \cdot S_n)/(1+R)^n = \phi_n \cdot \tilde{S}_n.$$

需注意的是它们都是适应过程，即 $V_n(\phi), \tilde{V}_n(\phi) \in \mathbb{R}^{\Omega_n}$.

一个策略 ϕ 称为是**自融资** (self financing) 的，是指

$$\phi_n \cdot S_n = \phi_{n+1} \cdot S_n, \qquad n = 0, 1, \ldots, N-1. \tag{8}$$

它的含义是：投资者在第 n 时刻为第 $n+1$ 时刻所作出的策略不改变其所掌握的证券组合的总价值，即投资者在此过程中既不需要增加资金，也不发生抽调资金。(8) 式显然等价于

$$\phi_{n+1} \cdot (S_{n+1} - S_n) = \phi_{n+1} \cdot S_{n+1} - \phi_n \cdot S_n,$$

从而

$$V_{n+1}(\phi) - V_n(\phi) = \phi_{n+1} \cdot (S_{n+1} - S_n) \quad \text{或} \quad \Delta V_{n+1}(\phi) = \phi_{n+1} \cdot \Delta S_{n+1},$$

其中 $\Delta V_{n+1}(\phi) = V_{n+1}(\phi) - V_n(\phi)$, $\Delta S_{n+1} = S_{n+1} - S_n$. 由此不难证明下列命题：

命题 8.1 策略 ϕ 自融资等价于下列条件之一:

i)

$$V_n(\phi) = V_0(\phi) + \sum_{j=1}^{n} \phi_j \cdot \Delta S_j, \qquad n = 1, \ldots, N;$$

ii)

$$\tilde{V}_n(\phi) = V_0(\phi) + \sum_{j=1}^{n} \phi_j \cdot \Delta \tilde{S}_j, \qquad n = 1, \ldots, N,$$

其中 $\Delta \tilde{S}_j$ 的定义类似。

在二期模型中，并没有什么自融资的概念。多期模型中之所以需要这个概念是为了使各期的投资活动有明确的联系。由命题 8.1 可看出，这种联系是指投资价值的改变都是由于证券价格改变所引起的，而与投资策略的改变无关。二期模型中不需要这个概念的原因在于那里只有一期投资策略，不存在投资策略改变的问题。同时，

自融资策略是指投资活动与外界财力无关，而不是意味着证券交易本身在同一时刻中不花钱。事实上，由于第 0 种证券实际上相当于一个可自由存取的银行帐户，其他证券交易所引起的财务不平衡都可由它来补足。这一事实其实已隐含在上述 ii) 中，因为 $\Delta\tilde{S}_j^0$ 总为零，对第 0 种无风险证券的策略 (ϕ_n^0) 其实并不出现在 ii) 的表达式中。更具体地说，自融资策略的实际维数只有 K 维，并且可表达为下列命题：

命题 8.2 对于任何风险证券组合策略 $((\phi_n^1,\ldots,\phi_n^K))_{n=0,1,\ldots,N}$ (可料过程) 和任何初值 V_0, 总存在唯一的无风险证券的策略 $(\phi_n^0)_{n=0,1,\ldots,N.}$ (可料过程), 使得策略 $\phi=(\phi^0,\phi^1,\ldots,\phi^K)$ 为自融资的，且有初值 V_0.

证明 ϕ_0^0 可通过下式计算得到：

$$V_0=\phi_0^0+\phi_0^1S_0^1+\cdots+\phi_0^KS_0^K.$$

为计算一般的 ϕ_n^0, 由自融资条件可得

$$\begin{aligned}\tilde{V}_n(\phi) &= \phi_n^0+\phi_n^1\tilde{S}_n^1+\cdots+\phi_n^d\tilde{S}_n^K\\ &= V_0+\sum_{j=1}^n(\phi_j^1\Delta\tilde{S}_j^1+\cdots+\phi_j^K\Delta\tilde{S}_j^K),\end{aligned}$$

从而 ϕ_n^0 就可由下式

$$\phi_n^0=V_0+\sum_{j=1}^n(\phi_j^1\Delta\tilde{S}_j^1+\cdots+\phi_j^K\Delta\tilde{S}_j^K)-\phi_n^1\tilde{S}_n^1-\cdots-\phi_n^K\tilde{S}_n^K$$

来唯一确定。 □

下面我们要对这样的市场提出无套利的概念。这里的“无套利”主要是指“未来值钱的资产当前也值钱”。也就是说，联系我们前面说过的“无套利假设的五个层次”，包括前四个层次的线性定价法则其实已经认为是“理所当然”的。从而在我们前面计算投资组合的价值时已经在“理所当然”地在那里运用。而“最高层次”的无套利假设将指出这一线性定价法则必须是“正”的。这种“正”性，现在将以所谓“等价概率鞅测度”的形式来表达。

首先我们称策略 ϕ 是**可接受的**，是指它是自融资的，且对于任何 $n=0,1,\ldots,N$ 满足 $V_n(\phi)\geqslant 0$. 称策略 ϕ 是**套利策略**，是指它是可接受策略，且满足 $V_0(\phi)=0, V_N(\phi)>0$, 其中后面的 $>$ 是按前面的向量意义来理解的，即 $V_N(\phi)$ 在任何状态下都非负，但至少在一个状态下为正。如果所讨论的市场不存在套利策略，那么该市场称为**可生存市场**。需注意的是我们在这里定义的套利策略要求它是可接受的。其实在这里去掉可接受的要求，对定义并无影响。事实上，我们可以证明下列命题：

命题 8.3　如果市场是可生存的，那么任何自融资策略都无法获得套利，即不存在自融资策略 (ϕ_n), 使得 $V_0(\phi)=0, V_N(\phi)>0$.

证明　事实上，如果存在某个自融资策略 ϕ, 使得 $V_0(\phi)=0, V_N(\phi)>0$, 那么也有

$$\tilde{V}_N(\phi)=\sum_{j=1}^{N}\left(\phi_j^1\Delta\tilde{S}_j^1+\cdots+\phi_j^K\Delta\tilde{S}_j^K\right)>0.$$

不妨设 ϕ 不是可接受策略。否则已由定义导得矛盾。令

$$n=\max\{k\,|\,\tilde{V}_k(\phi)\not\geqslant 0\,\}.$$

则设 $A_n=\{\omega\in\Omega\,|\,\tilde{V}_n(\phi)(\omega)<0\}$, 并定义新的策略 (ψ_n) 为

$$\psi_k(\omega)=\begin{cases}0, & k\leqslant n \text{ 或 } \omega\notin A_n,\\ \phi_k(\omega), & k>n \text{ 且 } \omega\in A_n.\end{cases}$$

那么不难验证，这也是个自融资策略，且

$$\tilde{V}_k(\psi)(\omega)=\begin{cases}0, & k\leqslant n \text{ 或 } \omega\notin A_n,\\ \tilde{V}_k(\phi)(\omega)-\tilde{V}_n(\phi)(\omega), & k>n \text{ 或 } \omega\in A_n.\end{cases}$$

由此可见，这是一个可接受策略，并且

$$\forall\omega\in A_n,\qquad \tilde{V}_N(\psi)(\omega)>0.$$

因此，与市场可生存矛盾。□

8.5　有了上面这些有关有限状态多期模型的基本概念以后，我们就可来讨论这一状况下的资产定价基本定理。我们在前面已经对有限状态二期模型指出资产定价基本定理，它断言，无套利假设等价于存在某种等价概率测度，使得买每一种证券的获利的数学期望值都等于

存银行的获利。在现在的有限状态多期的模型下，加上投资策略的自融资限制，更一般的资产定价基本定理将指出，这一结论不但要对最后时刻成立，还要对中间时刻也成立。为此，我们需要**鞅** (martingale) 的概念。

鞅的概念是 1939 年由法国数学家 Ville 首先引进的。鞅的字面意思是马颌缰，但其法文原意还有公平赌博的含义。有一种说法认为，“鞅 (martingale)”起源于法国的一个赌城的名字。在我们上述的框架中，鞅 (M_n) 被定义为一个适应过程，但是鞅的概念还必须与概率测度相联系，因为它与数学期望和条件数学期望的概念有关。适应过程 (M_n) 称为鞅是指它满足：

$$E[M_{n+1}|\mathcal{F}_n] = M_n, \qquad n = 0, 1, \ldots, N-1.$$

如果等号改为 $\leqslant$ ($\geqslant$), 那么 (M_n) 称为**上鞅** (**下鞅**)。不难看出，(M_n) 为鞅的充要条件也可表达为：

$$E[M_{n+j}|\mathcal{F}_n] = M_n, \qquad n = 0, 1, \ldots, N-1;\ j = 0, 1, \ldots, N-n;$$

尤其是由此可导出

$$E[M_n] = E[M_0] = M_0, \qquad n = 0, 1, \ldots, N;$$

即鞅在任何时刻的数学期望值都与初值一样。

我们以后的作为资产定价基本定理新形式的主要结论之一是：市场可生存的充要条件为存在等价概率测度，使得折现价格过程为鞅。因此，如果我们在下面有关鞅的一些结论中，都把鞅理解为折现价格，那么它们的经济意义就十分清楚。尤其是命题中的过程都可理解为多维的。

命题 8.4 设 (M_n) 是鞅 (折现价格)，(H_n) 是可料过程 (策略)，$\Delta M_n = M_n - M_{n-1}$. 那么如下定义的适应过程 (折现价值) (X_n) 也是鞅:

$$\begin{aligned} X_0 &= H_0 \cdot M_0 \\ X_n &= H_0 \cdot M_0 + H_1 \cdot \Delta M_1 + \cdots + H_n \cdot \Delta M_n, \quad n = 1, \ldots, N. \end{aligned}$$

证明 显然 (X_n) 是适应过程。同时，对于任何 n, 我们有

$$\begin{aligned} E[X_{n+1} - X_n|\mathcal{F}_n] &= E[H_{n+1} \cdot (M_{n+1} - M_n)|\mathcal{F}_n] \\ &= H_{n+1} \cdot E[M_{n+1} - M_n|\mathcal{F}_n] = 0, \end{aligned}$$

这里 H_{n+1} 可提出来是因为它是 $\mathcal{F}_n$-可测的。因此，

$$E[X_{n+1}|\mathcal{F}_n] = E[X_n|\mathcal{F}_n] = X_n,$$

即 (X_n) 是鞅。 □

这一命题意味着：只要价格是鞅，不管投资者采取什么 (自融资) 策略，其手中的证券组合的值也一定是鞅，即它的数学期望值总是不变的。下一命题说明，这一结论甚至还是价格为鞅的充要条件。

命题 8.5 适应过程 (折现价格) (M_n) 是鞅的充要条件为对于任何可料过程 (策略) (H_n), 有

$$E\left[\sum_{n=1}^{N} H_n \cdot \Delta M_n\right] = 0. \tag{9}$$

证明 如果 (M_n) 是鞅，那么 (9) 由上述命题可得。反之，为证明由 (9) 可导得 (M_n) 是鞅，我们来构造特殊的 (H_n). 事实上，为证明 $E[M_{n+1}|\mathcal{F}_n] = M_n$, 把随机变量理解为向量，只需指出，

$$\forall \omega_n \in \Omega_n, \qquad \frac{\sum_{\omega \in \omega_n} p(\omega) M_{n+1}(\omega)}{\sum_{\omega \in \omega_n} p(\omega)} = M_n(\omega_n). \tag{10}$$

因此，我们对固定的 $\omega_n \in \Omega_n$, 取

$$\begin{cases} H_k(\omega) = 0, & k \neq n+1, \\ H_{n+1}(\omega) = 1, & \omega \in \omega_n, \\ H_{n+1}(\omega) = 0, & \omega \notin \omega_n. \end{cases}$$

则对这样的 (H_n), (9) 立即导出 (10). □

现在我们来证明下面的重要定理：

定理 8.1 市场可生存的充要条件为存在等价概率测度，使得所有证券的折现价格过程为鞅。

证明 必要性。如果存在等价概率测度使得所有折现价格过程都为鞅，那么对于任何自融资策略 (ϕ_n), 其相应的证券组合的折现价值为

$$\tilde{V}_n(\phi) = V_0(\phi) + \sum_{j=1}^{n} \phi_j \cdot \Delta \tilde{S}_j.$$

由命题 3.3, $(\tilde{V}_n(\phi))$ 对该概率测度也是鞅，从而

$$E\left[\tilde{V}_N(\phi)\right] = \tilde{V}_0(\phi).$$

如果 (ϕ_n) 是可接受策略，那么当 $\tilde{V}_0(\phi) = 0$ 时，导出 $\tilde{V}_N(\phi) \geqslant 0$ 以及 $E[\tilde{V}_N(\phi)] = 0$. 这仅当 $\tilde{V}_N(\phi) = 0$ 时才有可能。因此，市场是可生存的。

充分性。在 $\mathbb{R}^\Omega$ 中考虑集合

$$M = \{\, x \in \mathbb{R}^\Omega \quad | \quad x = \tilde{V}_N(\phi) = \sum_{j=1}^{N} (\phi_j^1 \Delta \tilde{S}_j^1 + \cdots + \phi_j^K \Delta \tilde{S}_j^K),$$

$$(\phi_n) \text{ 为任何使 } V_0(\phi) = 0 \text{ 的自融资策略。} \}$$

和集合

$$K = \{\, y \in \mathbb{R}^\Omega \,|\, y > 0 \,\}.$$

那么它们是两个不相交的凸集，从而由凸集分离定理存在 $\mu = (\mu(\omega))_{\omega\in\Omega} \in \mathbb{R}^\Omega$ 使得

$$\forall x \in M,\ \forall y \in K, \qquad \mu \cdot y \geqslant \mu \cdot x;$$

同时，还有

$$\forall x \in M,\ \forall y \gg 0, \qquad \mu \cdot y > \mu \cdot x.$$

但由于 M 实际上是 $\mathbb{R}^\Omega$ 中的子空间，上两式的右端都只可能为零，否则它们都可能无限增大，以至不等式不可能成立。另一方面，由于 y 的每一分量都可无限增大，以至 μ 的每一分量都必须为正，否则上述严格不等式不可能成立。因此，$\mu \in \mathbb{R}^\Omega_{++}$, 从而可由它来定义概率测度为

$$\forall \omega \in \Omega, \qquad p(\omega) = \frac{\mu(\omega)}{\sum_{\omega'\in\Omega} \mu(\omega')}.$$

对于这一概率测度，由于对于任何 $x \in M$, 都有 $\mu \cdot x = 0$, 故可导得对于任何 K 维可料过程 $(\phi_n) = (\phi_n^1, \ldots, \phi_n^K)$, 有

$$E\left[\sum_{j=1}^{N} \phi_j \cdot \Delta \tilde{S}_j\right] = 0.$$

由命题 8.5, $(\tilde{S}_n)$ (及其每一分量) 是鞅。 □

与前面的讨论一样，可生存市场所导出的使所有证券折现价格为鞅的等价概率测度不一定是唯一的。要使这样的等价概率测度是唯一的，当且仅当上述证明中的子空间 M 的维数为 $\Omega - 1$. M 仅仅是所有初值为零的各种自融资策略所能达到的终值的全体。如果它的维数为 $\Omega - 1$, 那么再联上初值变化所形成的一维空间，它们就能张成整个 $\mathbb{R}^\Omega$. 回想起 $\mathbb{R}^\Omega$ 中的任何元素都能被称为衍生证券或未定权益，以及前面关于完全市场的定义，我们在此同样可以定义完全市场的概念。但是

为了在经济上更有意义，我们将在这里把未定权益限制为非负随机变量，并对完全市场的概念加以更多的要求。

一个 (非负) 未定权益 $h \in \mathbb{R}_{+}^{\Omega}$ 称为是**可达的**，是指存在一个可接受策略 ϕ, 使得 $V_N(\phi) = h$. 一个市场称为是**完全市场**，是指任何 (非负) 未定权益是可达的。这个定义在表面上看来增加了策略可接受的要求，其实对于可生存市场来说，这个要求自然成立。这是因为对于可生存市场来说，如果对于非负未定权益 h, 存在自融资策略 ϕ, 使得 $V_N(\phi) = h \geqslant 0$, 那么由于这时存在等价概率测度，使得 $(\tilde{V}_n(\phi))$ 为鞅，必然也有 $\tilde{V}_n(\phi) = E[\tilde{V}_N(\phi)|\mathcal{F}_n] \geqslant 0$, 即 ϕ 是可接受的。

下面我们给出本讲的另一条基本定理及其直接证明。

定理 8.2 可生存市场完全的充要条件为存在唯一的等价概率测度，使得所有证券的折现价格过程为鞅。

证明 必要性。假设市场是完全的。那么对于任何非负未定权益 $h \in \mathbb{R}_{+}^{\Omega}$, 存在可接受策略 ϕ, 使得 $h = V_N(\phi)$, 其中 ϕ 是可接受策略。如果存在两个等价概率测度 p_1 和 p_2, 使得 $(\tilde{V}_n(\phi))$ 为鞅，那么由这两个概率测度所定义的数学期望 E_1 和 E_2 将都有

$$E_i[\tilde{V}_N(\phi)] = E_i\left[h/S_N^0\right] = V_0(\phi), \quad i = 1, 2.$$

由于 h 是任意的，这仅当 $p_1 = p_2$ 时才有可能。

充分性。如果可生存市场不完全，那么由形为 $V_0 + \sum_{n=1}^{N} \phi_n \cdot \Delta\tilde{S}_n$ 的随机变量全体张成的空间 $\mathbb{R}^{\Omega}$ 的线性子空间 $\mathcal{V}$ 是真子空间。假定 p_1 是某等价概率测度，使得所有证券价格过程为鞅。那么由 p_1 生成的数学期望 E_1 可用来对 $\mathbb{R}^{\Omega}$ 定义内积：

$$\forall x, y \in \mathbb{R}^{\Omega}, \qquad (x, y) = E_1[xy].$$

由于 $\mathcal{V}$ 是真子空间，对于这个内积，存在 $z \in \mathbb{R}^{\Omega}$, 使得 z 与 $\mathcal{V}$ 正交，尤其是 $E_1(z) = 0$, 因为处处为 1 的随机变量是 $\mathcal{V}$ 的元素。定义新的概率测度

$$\forall \omega \in \Omega, \quad p_2(\omega) = \left(1 + \frac{z(\omega)}{2\max_{\omega\in\Omega}|z(\omega)|}\right) p_1(\omega).$$

那么它显然是一个与 p_1 不同的等价概率测度，并且由它所定义的数学期望 E_2, 将有

$$E_2\left[\sum_{n=1}^{N} \phi_n \cdot \Delta\tilde{S}_n\right]$$

$$= E_1\left[\sum_{n=1}^{N}\phi_n\cdot\Delta\tilde{S}_n\right]+\frac{1}{2\max_{\omega\in\Omega}|z(\omega)|}E_1\left[z\left(\sum_{n=1}^{N}\phi_n\cdot\Delta\tilde{S}_n\right)\right]=0.$$

由于上式对任何适应过程 $((\phi_n^1,\dots,\phi_n^d))$ 都成立，再由命题 3.5, 即得 $(\tilde{S}_n)$ 对于 p_2 也是鞅，即唯一性不满足。 □

至此，我们就完成了这个模型的整个定价理论框架。这就是说，对于一个完全市场来说，存在唯一的等价概率测度使得所有证券折现价格过程是鞅。而每个未定权益 h 又都可以用某个可接受策略 ϕ 来达到，使得 $V_N(\phi)=h$. 因为折现价值 $(\tilde{V}_n)$ 对于这一概率测度也是鞅，从而 h 在当前的定价就可确定为 $E[h/(1+R)^N]$. 这样我们就可对任何未定权益定价。 Cox–Ross–Rubinstein 模型只是这里讨论的模型的最简单的情形。

8.6 在上面的讨论中，由于数学上的困难，我们需要假定所有未来状态的个数都是有限的，从而所涉及的所有随机变量其实都是离散随机变量，所涉及的所有概率状态空间和 σ-域都是有限集。为讨论更一般的证券市场多期模型，我们首先需要去除有限状态的假定。这时，我们同样有“事件树”和 σ-域流的概念，但是它们的形象已经不是一棵枝杆分明的树，而要想象为一个有密密麻麻枝杈的“花菜”。

我们从一个概率空间 $(\Omega,\mathcal{F},\mathbf{P})$ 出发，这里 Ω 是一个任意的集合，$\mathcal{F}$ 是由 Ω 中的子集所构成的 σ 域。相应的 σ-域流仍可表示为 $(\mathcal{F}_n)_{n=0,1,\dots,N.}$, 这里每个 $\mathcal{F}_n$ 都是 Ω 的子集所构成的 σ-域，并且满足单调性：

$$0<m<n<N \implies \mathcal{F}_0\subset\mathcal{F}_m\subset\mathcal{F}_n\subset\mathcal{F}_N=\mathcal{F}.$$

这一定义实际上只与可测空间 $(\Omega,\mathcal{F})$ 有关，而与其上定义的概率 $\mathbf{P}$ 无关。正如我们在前面对于有限状态情形讨论的那样，σ-域流的定义与概率是无关的。但是在状态空间 Ω 是无限集时，为了避免一些不必要的麻烦，通常还对 σ-域流要求 Ω 中的概率为零的子集全体属于所有 $\mathcal{F}_n$. 也就是说，在所有由 $\mathcal{F}_n$ 所形成的各个概率空间中，零概率事件是相同的。因此，$\mathcal{F}_0$ 也就不再只有两个元素 $\{\emptyset,\Omega\}$ (代表“肯定不发生事件”与“肯定发生事件”)，并且还包含所有概率为 1 的事件和所有概率为零的事件。不过，我们经常把肯定不发生事件与零概率 (“几乎肯定”) 事件不加区别。

随机变量的概念其实也与概率测度无关。所谓可测空间 $(\Omega,\mathcal{F})$ 上的随机变量 X 是指它是定义在 Ω 上的可测函数，即对于任何实数 c, $\{\omega\in\Omega\,|\,X(\omega)<a\}\subset\mathcal{F}$, 或者说 $\{X<c\}$ 是一个事件。把 $\mathcal{F}$ 替代为它的子 σ-域 $\mathcal{F}_n$, 我们就可得到相应的 $\mathcal{F}_n$-可测随机变量概念。若干个随机变量按次序放在一起考虑仍然称为随机序列随机过程。在我们考虑的问

题中，随机过程总是有下列形式：

$$x = (x_n)_{n=0,1,\ldots,N.}$$

这里每个 x_n 同样都可以是多维随机变量，(x_n) 可以是多维随机过程。适应过程 (S_n) 定义为它在第 n 时刻时的值 S_n 是 $\mathcal{F}_n$-可测的。可料过程 (ϕ_n) 同样定义为，除了当前外，它在第 n 时刻时的值 ϕ_n 是 $\mathcal{F}_{n-1}$-可测的。在这种情形下，鞅 (M_n) 也可类似定义，即它是适应过程，并且满足

$$E[M_{n+1}|\mathcal{F}_n] = M_n, \qquad n = 0, 1, \ldots, N-1.$$

前面讨论的结果在目前情形下都仍然成立，只是所有的等式都要理解为“几乎肯定 (almost sure, a.s.)”成立，即以概率 1 成立。事实上，我们上面的讨论中需要用到状态有限的地方是在资产定价基本定理 8.1 的必要性证明中，其他的证明都不需要本质的改变。定理 8.1 的必要性证明中主要用到凸集分离定理。这一定理在无限维的情况下将受到一定的限制。但是 1990 年，Dalang–Morton–Willinger [7] 证明，这条定理对于状态无限的概率空间仍然成立，其中用到一条相当于凸集分离定理的证明相当细腻的概率论文本。而其中涉及的等价概率测度则是指使零概率事件不变的概率测度，或者说是对原来的概率测度绝对连续的概率测度。然而，如果模型的期数 N 也是无限时，那么类似的资产定价基本定理就不再成立。其反例可在 [7] 或 Pliska [18] 中找到。

至今，我们为导出 Black–Scholes 期权定价公式以及一般的期权定价原理而讨论了证券市场的多期模型。其中着眼点主要在于一般的资产定价基本定理，而对隐含假设的线性定价法则完全忽视。回想起我们在前五讲中，线性定价法则曾使我们获得资本资产定价模型、Markowitz 证券组合选择理论等许多重要结果，一个自然的问题是：线性定价法则以至随机折现因子理论是否对证券市场多期模型也能导出类似的结果？这个问题的答案应该是肯定的，但有关的研究似乎还不很充分。而一些初步的讨论，已经能使我们对经典的理论有更深入的理解。事实上，实际的证券市场并非是二期的，而是多期的，甚至是“连续期”的。二期模型使时间因素凝固化，而多期模型才能突出时间因素所起的重要条件作用。

现在我们可以对上述的 σ-域流来定义“未定权益 Hilbert 空间”结构。基本的“未定权益 Hilbert 空间” $\mathcal{M}$ 当然定义为概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 上的一些方差有限 (平方可积) 的随机变量全体，它的经济含义是 N 天后的未定权益的价值随机变量。它们的当前价值仍然由某个定义在 $\mathcal{M}$ 上的定价函数 p 来确定。如果这个 p 线性连续，那么存在 $m \in \mathcal{M}$, 使得对于任何 $x \in \mathcal{M}$, 有 $p(x) = E[mx]$. 现在的问题是我们还有一些中间空间要考虑。这些中间空间是指 $\mathcal{M}$ 中的 $\mathcal{F}_n$-可测的全体 $\mathcal{M}_n$ 所形成一

些 $\mathcal{M}$ 的子空间。可以证明，这些子空间都是闭的，即它们都是 Hilbert 空间。于是我们有

$$\mathbb{R} = \mathcal{M}_0 \subset \mathcal{M}_1 \subset \cdots \subset \mathcal{M}_n \subset \mathcal{M}_{n+1} \subset \cdots \subset \mathcal{M}_N = \mathcal{M}.$$

在这个框架中讨论的“未定权益适应过程” (x_n) 可定义为满足 $x_n \in \mathcal{M}_n$ $(n = 0, 1, \ldots, N)$ 的随机过程，其含义是未定权益在各个时刻的值。于是我们不但要考虑定义在 $\mathcal{M} = \mathcal{M}_N$ 上，取值在 $\mathcal{M}_0 = \mathbb{R}$ 上的定价函数，还要考虑对于 $0 < s < t \leqslant N$, 定义在 $\mathcal{M}_t$ 上取值在 $\mathcal{M}_s$ 上的定价函数。它的含义是：时刻 n 的未定权益在时刻 m 所取的 (不确定) 值。如果这样的定价函数也是线性连续的，那么它是否也有类似的“随机折现因子”存在？这个问题作为数学问题，需要推广经典的 Riesz 表示定理。而这正是 Hansen–Richard [10] 为了讨论条件信息的作用的主要结果。在我们的框架中，这一结果可叙述如下：*设 $0 \leqslant s < t \leqslant N$. 那么 $p_s^t\colon \mathcal{M}_t \to \mathcal{M}_s$ 为线性连续算子的充要条件为: 存在唯一的 $m_s^t \in \mathcal{M}_t$, 使得*

$$\forall x_t \in \mathcal{M}_t, \qquad p_s^t(x_t) = E[m_s^t x_t | \mathcal{F}_s].$$

这样一来，对于一个实际的金融问题，如果我们考虑未来的两个时刻：t 和 $t+1$, 我们可以讨论这样的随机折现因子理论：

$$p_t(x_{t+1}) = E[m_t^{t+1} x_{t+1} | \mathcal{F}_t] = E_t[m_t^{t+1} x_{t+1}],$$

这里 x_{t+1} 是 $t+1$ 时刻的 ($\mathcal{F}_{t+1}$-可测) 未定权益， E_t 表示对 $\mathcal{F}_t$ 的条件数学期望。这个等式与我们在讨论二期模型中的根本不同在于它将体现时间因素的条件信息作用，并且定价函数的取值也是不确定的随机变量。与此相联系的，由此出发，我们同样可以得到资本资产定价模型和 Markowitz 均值－方差前沿，但是它们都将是带条件信息的随机模型。这样，我们将更好地理解经典金融经济学实证分析。历史上，曾经对资本资产定价模型的实证检验有过长时期的争论。但是如果考虑到条件信息，这些争论是有可能得到一个合理的解释的。

事实上，所谓“条件数学期望”也是一个随机变量，它可以理解为“在不同条件 (状态) 下，取不同值的”的量，但对它很难直接作统计分析。然而，从数学上来说，上述等式等价于：对于任何 $\mathcal{F}_t$-可测的随机变量 z_t, 有 $E[z_t p_t] = E[m_{t+1} x_{t+1} z_t]$. 在这最后一个等式中，数学期望已经是无条件的，从而就有可能来进行统计计算。因此，为测算条件数学期望，我们可以通过引进若干 z_t 来进行。这种 z_t 称为**工具变量**。目前有许多研究条件信息的金融实证分析都是通过这种方法来进行的。例如，我们曾经在“代引言”中提到过的 Fama–French 的三因子资产定价公式，它在市场组合以外，再引入公司规模以及公司的帐面值与市值

之比作为因子，就可解释为对“条件资本资产定价模型”的统计检验，而不是对资本资产定价模型的否定。关于这方面的进一步讨论可参看 Cochrane [4]。

还有一个问题：在完整的无套利假设下，前面已经指出，它等价于对于所有基本证券 (包括无风险证券) 的折现价格过程，存在等价概率鞅测度。把这一结果与随机折现因子理论相比较，我们可以看到，在 $p(x)=E[mx]$ 中的随机折现因子 m 的“增值” $m(1+R)^N$ 相当于等价概率鞅测度对原概率测度的所谓 Radon–Nikodym 导数。不但如此，由于这个等价概率测度并不依赖于时期，故对于 t $(>s)$ 时期的未定权益价值 x_t 的折现与它在 s 时期的价值定价$p_s^t(x_t)$ 的折现之间应该有下列等式：

$$\frac{p_s^t(x_t)}{(1+R)^s}=E\left[m(1+R)^N\frac{x_t}{(1+R)^t}\right].$$

与上述 Hansen–Richard 所推广的 Riesz 表示定理相比较，我们得到

$$p_s^t(x_t)=E[m_s^t x_t|\mathcal{F}_s]=E[m(1+R)^{N-(t-s)}x_t|\mathcal{F}_s].$$

这就是说，在一定意义下，随机折现因子随时间的变化相当于同一个随机折现因子的“增值”变化。由此导得的“时变”资本资产定价模型因而也就变为简单的折现关系。但是这是通过完整的无套利假设成立而得到的。类似的论证是否可直接通过多期模型的线性定价法则来得到似乎还不得知。然而，从实证的观点来看，我们或许更希望有一个资本资产定价模型如何更复杂地随时间变化的模型。这种模型的一个文本是所谓“跨时资本资产定价模型 (Intertemporal Capital Asset Pricing Model, ICAPM)”。后者是 Merton [17] 从连续时间金融学的思路出发总体考虑时间的结果，其最终形式如同经典的 CAPM 随时间变化再加上一项对状态变化的考虑。它可为某些多因子模型提供依据。

说明 自从 Black–Scholes 期权定价公式 ([2]) 于 1973 年问世以来，确确实实对金融经济学和金融业掀起一场革命。但是 Black–Scholes 公式也并非是从天上掉下来的。我们从 8.2 节中所引述的一些历史上曾经出现过的一些公式中，可以看出，这一公式的出现决不是偶然的，而是经过许多学者前赴后继的努力才达到的成就。有关历史材料转引自 Briys–Bellalah–Mai–de Varenne [3], 在那里还有更详尽的介绍。Black–Scholes 的经典论文中也同样提到一些前人的工作。

有关 Black–Scholes 理论的专著和教材目前几乎可以用“多如牛毛”来形容。对于商学院的学生来说，标准的教材是 Hull [13]，但是所有金融经济学方面的著作 [8], [9], [12], [14], [15] 都必须有 Black–Scholes 理论的陈述。由于数学上的困难， [12] 和 [14] 都没有涉足很深，而停留在简单的二叉树方法。[8], [9] 和 [15] 则都讨论得比较深入。不过 Huang–Litzenberger [12] 的 162–166 页在适当的假设下，对一个二期模型中有常相对风险厌恶效用函数的个体投资者的投资决策问题也导出了 Black–Scholes 公式。这在机理上对人们有新的启示。

我们的讲义的主要内容在很大程度上来自 Lamberton–Lapeyre [16], 该书的特点在于数学上条理比较清楚。如果有读者希望对 Black–Scholes 理论的数学上的每一细节都不放过，但又不愿意陷入过深的具体证明，那么可以建议他读 Baxter–Rennie [1]，该书相当通俗地介绍了所有有关的数学，但是几乎没有一个严格的证明。关于离散时间模型的期权定价理论讨论的另一本标准专著是 Pliska [18]。但是由于该书作者坚持“离散到底”，居然连 Black–Scholes 公式都没有导出。这不能不说是一个缺陷。

Black–Scholes 原来的推导是建立在连续模型的基础上的，有关论述我们将在以后进行。用二叉树方法给出期权定价的初等证明要归功于 Cox–Ross–Rubinstein [5] 以及 Rendleman–Bartter [19]。这一证明使人们更容易接受 Black–Scholes 公式的机理及其与资产定价基本定理之间的关系，并且它本身也为 Black–Scholes 理论提供了一种计算方法。Cox–Rubinstein [6] 更是进一步把这样的思想展开成一本专著，它是期权定价理论中少有的一本不涉及随机分析、但又有透彻的数学证明的著作。对于见到随机分析“发怵”的读者来说，可能读这本书会感到轻松些。

资产定价基本定理对于多期模型有了新的形式，其中最重要的是引进了鞅的概念。但是我们的讨论仍限于有限状态情形。对于一般情形我们可以想象其数学上的复杂性。这正是近年来许多金融数学家在不断精益求精地研究的问题。在最后一节中，我们介绍了一般的多期模型中的资产定价基本定理。有限时期、有限基本证券个数可能已经是“无套利假设与存在等价概率鞅测度等价”的极限。时期无限 (尤其包括连续时间情形) 或基本证券个数无限都将使得资产定价基本定理需要修正。

把完整的无套利假设减弱到线性定价法则来讨论多期模型应该说是一种很自然的想法。尽管已经有了 Hansen–Richard [10] 的重要定理，使得随机折现因子理论可适应多期模型。但是到目前为止，这方面的理论研究似乎还不充分。

思考题和习题

1. 什么是 Black–Scholes 期权定价公式，其中的每一个参数是什么意义？

2. 试利用计算器，计算若干组欧式期权的定价。

3. 由 Black–Scholes 公式出发，可定义许多灵敏度参数。这些参数通常用希腊字母表示。对于欧式买入期权来说，这些参数有

$$\Delta = \frac{\partial c}{\partial S}, \quad \Gamma = \frac{\partial^2 c}{\partial S^2}, \quad \Theta = \frac{\partial c}{\partial T}, \quad \mathrm{V} = \frac{\partial c}{\partial \sigma}, \quad \mathrm{P} = \frac{\partial c}{\partial r}, \quad \Omega = \frac{S}{c}\frac{\partial c}{\partial S},$$

其中 V 并不是希腊字母，但它被称为 vega (织女星)，有时它也被表示为 Λ (lambda), K (kappa)， P 是希腊字母 rho 的大写。试根据 Black–Scholes 公式，求出这些“希腊字母”的表达式，并根据偏导数的意义来指出这些“希腊字母”的作用。

4. 对于欧式买入期权来说，其定价一般当然是执行价格的递减函数，即执行价格越高，期权越便宜。但是从 Black–Scholes 公式来看，是否一定有 $\partial c/\partial K < 0$?

5. 怎样理解卖出－买入期权平价关系 (6)? 它的极限情形是怎样的关系？试指出，如果这个关系不成立，那么就可构造一个套利组合 (可对 $N = 1$ 先讨论)。

6. 从 Black–Scholes 公式的二叉树方法推导中，我们假设在每一时刻每一状态下股价可能的变化都是按同样的比例升降。这在股市中究竟意味着一种怎样的假设？

7. 什么是“事件树 (信息流)”？怎样来理解它意味着“事件信息越来越清晰”的过程？数学上它是怎样用向量空间来表示的？

8. 什么是 σ-域流 $\{\mathcal{F}_n\}$? 什么是 $\mathcal{F}_n$-可测的随机变量？什么是随机变量的条件数学期望？

9. 什么是适应过程？什么是可料过程？怎样用证券价格过程和投资策略过程作为模型来理解这两个数学概念？

10. 什么是自融资策略？对于自融资策略 $\{\phi_n\}$, 怎样理解 $\Delta V_{n+1}(\phi) = \phi_{n+1} \cdot \Delta S_{n+1}$?

11. 什么是可接受策略？什么是套利策略？什么是可生存市场？

12. 怎样定义一个鞅？对于金融经济学来说，鞅的概念为什么非常重要？

13. 对于多期模型来说，怎样用鞅来表达资产定价基本定理？这一定理的条件是什么？

14. 怎样把随机折现因子理论与多期模型联系起来？

参考文献

[1] Baxter, M., and Rennie, 1996, A., Financial Calculus, An Introduction to Derivative Pricing, Cambridge University Press, Cambridge.

[2] Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, **81:** 637–654.

[3] Briys, E., M. Bellalah, H. M. Mai and F. de Varenne, 1998, Options, Futures and Exotic Derivatives: Theory, Appplication and Practice, John Wiley & Sons, Chichester-New York (中译本：埃里克·布里斯等，2002，期权、期货和特种衍生证券：理论、应用和实践，史树中等译，机械工业出版社).

[4] Cochrane, J. H., 2001, Asset Pricing, Princeton University Press, Princeton, N. J..

[5] Cox, J., S. Ross and M. Rubinstein, 1979, Option pricing: A simplified approach, *Journal of Financial Economics,* 3: 145–166.

[6] Cox, J., and M. Rubinstein, Option Markets, 1985, Prentice Hall, New Jersey.

[7] Dalang, R., A. Morton & W. Willinger, 1990, Equivalent martingale measures and no-arbitrage in stochastic securities market models, *Stochastics and Stochastic Reports*, **29**: 185–201.

[8] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[9] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N. J..

[10] Hansen, L. P. & S. F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica*, **55:** 587–614.

[11] Harrison, J. M., and D. M. Kreps, 1979, Martingales and arbitrage in multiperiod securities market, *Journal of Economic Theory*, 20: 318–408.

[12] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[13] Hull, J. C., 1999, Options, Futures, and Other Derivatives, 4th ed., Prentice–Hall, N.J.. (第三版有中译本：赫尔，J., 2000，期权、期货和衍生证券，张陶伟译，华夏出版社).

[14] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[15] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[16] Lamberton, D., et B. Lapeyre, 1991, Introduction au Calcul Stochastique Appliqué à la Finance, Ellipses, Paris (英译本：Lamberton, D., & B. Lapeyre, 1996, Introduction to Stochastic Applied to Finance, Chapman & Hall, London).

[17] Merton, R. C., 1992, Continuous-Time Finance, Rev. ed., Blackwell, Oxford.

[18] Pliska, S. R., 1997, Introduction to Mathematical Finance: Discrete Time Models, Blackwell, Massachusetts (中译本：斯坦利 • R • 普利斯卡，2002，数理金融学引论：离散时间模型，王忠玉译，经济科学出版社).

[19] Rendleman, J. R. Jr. and B. J. Bartter, 1979, Two-state option pricing, *Journal of Finance*, 34: 1093–1110.

九、有效市场理论

本讲要求：了解有效市场概念的基本思想及其确切内涵。了解三种形式的有效市场的定义。理解信息集的一种定义以及理性预期均衡的概念。

数学预备知识：概率论中的随机游走概念和 σ-域的概念在有效市场理论中起本质作用。

内容：

- 9.1 有效市场的通俗理解和讨论。
- 9.2 有效市场假设的历史回顾。
- 9.3 有效市场的检验。
- 9.4 信息集的一种定义以及理性预期均衡。
- 9.5 从理性预期均衡来看 CAPM 和 APT。

我们在前面几讲中已经看到，对于市场的无套利假设在金融经济学中起着核心作用。那么人们自然要问，金融市场是否符合无套利假设？这个假设并不像其在普通商品市场中的对应物“无投入就无产出”那样无可争辩。对于普通商品来说，“无中生有”只能是幻想。但是对于金融市场来说，“空手套白狼”的事例似乎时有听说。因此，一个市场是否符合无套利假设，在什么条件下符合无套利假设就值得研究。这就是本讲的主题：有效市场理论。

9.1　整个有效市场理论的思路不难理解。市场上的套利机会就像大街上撒的钱。如果不考虑法律、道德之类的约束，谁都可以去捡，那么随着这个消息被不断透露出去，这样的钱很快就会被人们捡完。因此，从总体来说，大街上是不会有钱等着人们去捡的。如果有人捡到了钱，那也是极个别的事。任何“发财秘诀”(其中包括投资家的经验之谈、股价预测、各种基本分析和技术分析方法等等) 都没有普遍意义。

这样的观点自然会引起许多争论。不但是从事基本面分析、技术分析的金融业从业人员不能接受，经济学家们也对此有怀疑。例如，著名的 Grossman–Stiglitz 悖论 ([8]) 实质上是把上述推理推向极端造成的：既然大家都认为大路上没有钱可捡，于是谁也不考虑到大路上去捡钱。这样，大路上真的有钱时，就谁也不去捡。然而，必然有聪明人想到这一情形，他就会去捡别人想不到去捡的钱。于是这个聪明人就不断可在大路上捡钱。但是既然有这样的聪明人存在，别人也不会都是傻子，这样的钱还是很快就都被人捡走了。如此等等就成了一个逻辑上无法自圆其说的悖论。正如我们在第一讲中已经提到，这一推

理过程中已经涉及高阶逻辑，通常的数学和逻辑工具在这里已经不够用。Grossman–Stiglitz 的解决办法是强调信息的搜集是有成本的。尽管如此，从“大街上无钱可捡”这个假设出发来看待金融市场理论仍然是非常重要的。

有效市场理论的着眼点是信息。它与两次“华尔街革命”所形成的金融经济学的理论框架没有直接联系，而属于理论前提是否成立的另一层次的研究。在方法论上与我们前面的讨论相比较也有很大不同。近年来，这方面的研究又进一步与信息经济学、行为科学等联系起来，正在使我们对整个金融市场的运动又有更深入的了解，而不再简单地停留在“无套利假设是否成立”上。即使对于“大街上无钱可捡”这一似乎是“显然”的命题，现在也有人提出质疑：如果在车水马龙的高速公路上，有一张 100 美元，谁都看到了，但谁也不敢贸然去捡。也就是说，“捡钱”也可能伴随着“风险”和其他条件，不一定有钱就一定有人捡。因此，如果发现“捡钱机会”，并不等于一定有人去捡。这就是说，即使存在套利机会，也不一定与市场有效矛盾[1]。行为金融学的研究者们更是在这点上大加发挥。在 Shleifer 的 2000 年的新著 [29] 中，第一章就题为“金融市场有效吗 (Are Financial Markets Efficient) ?”。其中尤其提出“非理性交易者”、“噪声交易者”等存在对市场的影响。这一章的讨论能使我们对有效市场理论有更清醒的认识。我们这一讲的目标仍在于介绍这方面的一些理论观念及其数学公理化表达形式。

9.2　关于有效市场理论的历史及其金融含义在各种《投资学》教科书中都有不少讨论。这里我们对此只作简单介绍。

通常认为，有效市场理论与期权定价理论一样，都起始于 Bachelier 1900 年的论文。这里主要指他对股价的认识，并且涉及所谓**随机游走** (random walk) 的概念。随机游走最简单的模型就是上一讲中提到的“二叉树”。这个二叉树模型也可被设想为一个在一条直路上横行的醉汉的运动。醉汉在每一时刻都可能往右走或者往左走。于是在 N 个时刻以后，他可能到达的地方有 2^N 种可能。如果用数学来刻画这样的运动，那么我们可以说这样的随机游走是一个随机序列，序列的每一项对应醉汉到达的坐标。这个序列有这样的特点：序列增量所对应的随机序列的每一项是独立同 (二项) 分布的随机变量 (或往左、或往右一步)。这样，序列本身将有这样的特点：它的每一项 (在任意时刻的位置随机变量) 的均值都等于其起点位置，而方差则与时间间隔成正比。如果我们再进一步联系上一讲中提出的鞅的概念，还可以知道，这个“随机游走序列”是一个鞅。

---

[1] 在 Lo–MacKinlay [22] 中，则举艾滋病疫苗为例。如果有这样一种疫苗，它肯定有巨大的获利机会。但是这一“捡钱”机会谁都看到了，却并不是谁都能捡到的。参看本讲的附录。

如果再把随机游走的概念连续化，即假定时间是连续变化的，“醉汉”也是在每一时刻都“左右摇摆”地乱走，那么它就变成所谓 **Brown 运动**的概念。Brown 运动原来是指苏格兰生物学家 R. Brown （1773–1858) 在显微镜底下发现的花粉颗粒的不规则运动。它在数学上可定义为一个随机过程 (即以时间为自变量、随机变量作为因变量的函数)，其样本轨线要求是时间的连续函数，同时它在不同时刻的小时间间隔的增量是独立同分布的随机变量。以前人们都以为 Brown 运动的数学定义是 Einstein (1880–1955) 首先于 1905 年提出的，并由此解决了一些统计物理问题。自从发现了 Bachelier 的学位论文以后，人们才注意到，第一个为 Brown 运动给出数学定义的是 Bachelier, 并且作为一位数学家，他在对于 Brown 运动数学研究的各方面都比 Einstein 要深入。不但如此， Bachelier 还在此基础上，历史上第一次给出一个期权定价公式 (见上一讲)。可惜的是他错误地假设股票的价格变化是 Brown 运动，而不是股价的收益率变化是 Brown 运动，以至最后得到的期权定价公式与后来的 Black–Scholes 公式相去甚远。然而，他的历史功绩是不可磨灭的。

Bachelier 以后再次提出股价变化是一种随机游走的经济学家是 A. Cowles [2] (1891–1984)。他在他 1933 年发表的一篇题为《股市预测家能预测吗？》的著名论文 [3] 中，通过实证数据分析认为，股价的变化是不可预测的。再以后，影响更大的就是 1965 年 Samuelson 在已经发现 Bachelier 的研究以后，发表了一篇题为《恰当预期价格随机涨落的证明》([28])，再次强化了价格的随机性，这里“恰当预期 (properly anticipated)”被理解为价格真正融合了所有市场参与者的期望与信息。**有效市场假设** (Efficient Markets Hypothesis, EMH) 的提法也始于 Samuleson。最后，完整地讨论有效市场假设的是 Fama 在 1965 年发表的题为《股市价格中的随机游走》的论文 ([6]) (正式发表时改名为《股市价格的行为》，见 [7]) 和 1970 年发表的综述论文 [8]。在后一篇论文中，他明确地提出：“价格总是 ‘完全反映’ 可接受信息的市场称为 ‘有效’ 的 (A market in which prices always ‘fully reflect’ available information is called ‘efficient’)”。在 Fama 的经典论文 [7], [8] 以及今天在许多《投资学》教科书中都提出的有效市场的分类，则是出于 Robert 1967 年没有公开发表的一次讨论会上的发言，它把有效市场从信息集的角度在形式上分为三类：

1. **弱有效**：信息集只包含价格或收益自身的历史 (公共信息)，它使技术分析无效；
2. **半强有效**：信息集包含所有对所有市场参与者都已知的信息 (公开信息)，它使基本分析无效；

---

[2] 与这位投资家兼经济学家的名字紧密相连的是 Cowles 基金会，这是个“盛产”诺贝尔经济学奖得主的研究机构。

3. **强有效**：信息集包含所有对某个市场参与者已知的信息 (私人信息)，它使一切“黑箱操作”无效。

许多教科书对这三种形式上的有效市场都作出非常具体生动的说明。但是这类说明多半都带有一定的“文学色彩”，而这种“文学色彩”实际上为理论添加不少夸张 (例如，声称“价格已经反映了海湾战争的影响”等等)。对于上面各种有效市场假设的提法的正确理解在于：人们不能再利用所述的信息集，得到额外的收益，而有关的非公开信息如果得到披露，也不会引起价格的波动。

这里我们有一些基本概念需要澄清。如上所说，随机游走是鞅，而资产定价基本定理则是指无套利假设在一定条件下 (尤其是在离散模型中，证券种类有限，时期数有限的情况下) 等价于存在一个等价概率测度，使得证券的折现价格过程是鞅 (从而证券的收益过程也形成鞅)。而对于客观概率来说，后者不一定是鞅。因此，如果能检验到 (折现) 股价是随机游走，那么无套利假设就成立。但是反过来不成立，即无套利假设成立并不要求股价是随机游走，甚至并不要求股价是鞅。在这方面学术界曾有过争论。 Fama 的有效市场定义在理论上并不明确 (这里我们要考虑到 Fama 的经典论文出现在 Black–Scholes 理论和 Ross 的资产定价基本定理以前，无套利假设的观念没有像后来那样清晰)，因此，曾受到 LeRoy [18] 的批评。Fama 作了回答 ([9])，但似乎并未说服 LeRoy ([19])。其实 Fama 最初看来是赞同随机游走假设的。后来实际上把有效性检验变为鞅性质检验。LeRoy 以及 Lucas [23] 也都认为市场有效性等同于股价的鞅性质。在这点上，Fama 与他们似乎并没有什么分歧。而分歧之处则在于对信息集的理解 (Fama 强调“市场信息”)。因此，所谓弱有效的检验通常就归结为鞅性质的检验。而 Ross [27] 则认为市场有效应该等同于无套利假设成立。这一见解虽然很有道理，但它是无法检验的。这也许为有效市场理论留出更大的空间。然而，目前对有效市场理论的突破更多的是对无套利假设的突破。因此， Ross 的见解就不起多大作用了。

此外，通常所说的随机游走还允许有“漂移”，即它可定义为 $x_{t+1} = \mu + x_t + \varepsilon_t$, 其中 ε_t 对于任何 t 是均值为零的随机变量，并且当 $t \neq t'$ 时 ε_t 与 $\varepsilon_{t'}$ 独立同分布。这样定义的随机游走当 $\mu \neq 0$ 时，它就不是鞅。因此，一般来说，随机游走假设既不是市场有效的充分条件，也不是市场有效的必要条件。甚至连鞅性质也只是在近似意义下可看作市场有效的充分条件。这是因为鞅性质无非意味着证券价格的不可预测性，而 Lucas [23] 指出，在有效的理性预期一般均衡下的市场中，资产价格有与可预测的消费有关的预测成分。因此，当 Lo 和 MacKinlay [21] 于 1988 年实证指出，股价即使对于广义的随机游走假设都不满足时，虽然也曾引起一些小小的轰动，但很快也就被人们所接受。总之，今天

人们对有效市场的认识早已越过随机游走或鞅的观念。

9.3　有效市场问题的研究通常都是通过实证检验来进行的。这方面的具体检验方法要涉及许多数理统计技巧，超出本讲义的范围。但是我们可以概述某些思想。

正如我们在前面所说，弱有效假设以前常被认为等价于证券价格或收益的随机游走假设或者鞅假设，或者后两者至少是必要条件。于是历史上曾经有过许多实证检验，来肯定或否定随机游走假设或者鞅假设。总体来说，对于美国证券市场来说，多数检验是肯定的。但是后来的研究发现，这两个假设在理论上对于弱有效假设来说并非必要。同时，从理论上来说，弱有效假设的检验应该是一个联合检验，即一方面它要对一个定义正常收益的市场的均衡模型进行检验，另一方面它再对收益检验它是否是正常的。这样一来，弱有效检验就变得复杂起来。如果有效性检验没有通过，那么这并不能说明市场一定是无效的，因为这还有可能是因为市场的均衡模型检验没有通过。如果随机游走检验或鞅检验通过，但是对市场的均衡模型检验 (即无法肯定收益是否正常) 没有通过，那么仍然不能肯定弱有效假设检验通过，即随机游走假设和鞅假设对于弱有效假设也不是充分的。原来认为，弱有效假设成立，技术分析就无效。现在大家开始认为，情况不一定就那么简单。Fama 的有关有效市场的第二篇经典综述 [10] 就不再说弱有效市场假设意味着价格不可预测。而一本近年来影响很大的专著 Campbell–Lo–Mackinlay [1] 的第二章的标题就是：资产收益的可预测性。在这一章的结语中，作者明确指出：“**近年来计量经济学的进展和实证证据看来在暗示金融资产收益在某种程度上是可预测的。如果在三十年前这将等于是说彻底否定了有效市场。然而，当代金融经济学告诫我们，对于这样的可预测性还有其他完全合理的因素可以考虑。证券市场的精细结构和交易过程中的摩擦可能生成可预测性。改变商情条件引起的时变期望收益可能生成可预测性。一定程度的可预测性对于回报承担动态风险的投资者来说将是必要的。**”这样的观点看来会越来越引起人们的重视。

关于半强有效假设和强有效假设的实证检验研究也很多。Fama 把这类研究称之为“事件研究 (event study)”。具体的方法就是看某一信息披露前后是否对价格有影响。我们应该注意到，这类检验对于“证伪”是有效的，即如果检验的结果是否定的，我们可以相当肯定地说，信息披露对价格有影响；然而，这类检验对于“证实”并非很有效，即如果检验的结果是肯定的，我们只能说，你所用到的信息披露，在你所检验的时间段中对价格影响不大，但不能说所有有关信息披露都对价格披露都没有影响。这种“证伪”与“证实”的不对称性倒并不是专对有效市场实证研究的。事实上，所有实证研究都存在这样的一个根

本问题。只要抓住一个白乌鸦，就能否定“天下乌鸦一般黑”，但是抓住一万个黑乌鸦，也只能加强我们对“天下乌鸦一般黑”的信念，却不能完全肯定“天下乌鸦一般黑”。归根结蒂在于经验归纳虽然对于建立科学理论、寻求一般规律是必要的，但仅仅只有经验归纳是不可能建立科学理论和一般规律的。

对于半强有效假设的研究肯定结果与否定结果几乎一样多。但是对于强有效假设的研究则否定结果比肯定结果要多。这就是说，利用内部信息，一般是能够获得超额收益的。因此，世界各国都有利用内部信息进行证券交易非法的法律条文规定，以保证证券市场的健康发展。具体的研究结果可参看 Fama [7], [8], [10] 以及许多《投资学》教科书中的介绍。

最近十几年来，出现了大量的否定有效市场假设的实证检验研究。这方面的典型文献，除了我们上面已经提到的以外，有例如，Cochrane [2]， Malkiel (他是 1973 年出版的名著鼓吹有效市场假设的《沿华尔街的随机游走 (A Random Walk Down Wall Street)》的作者。在那本书中，作者夸张地提出根据一头蒙着眼的黑猩猩投飞镖来选择投资组合能与专家的选择一样好。Lo–Mackinlay [22] 的书名就是要与他唱对台戏) [25], Shiller (他是最近曾经访华的行为金融学奠基人之一。其 2000 年出版的名著《非理性繁荣 (Irrational Exuberance)》已经有中文版) [30] 等。在 Cochrane [2] 的引言中，作者几乎以“忏悔”的口吻来简要生动地叙述了这一历史转变：“最近 15 年来，人们已经看到金融经济学家在理解投资界的方式上有一场革命。我们曾经以为股票和债券的收益本质上是不可预测的。而今我们承认股票和债券的收益长期来说有一个可观的可预测成分。我们曾经以为资本资产定价模型 (CAPM) 对于为什么某些股票、组合、基金或投资策略的平均收益比别的高提供了一种很好的描述。而今我们承认许多投资机会的平均收益不可能用 CAPM 来解释，“多因子模型”用来取代它的位置。我们曾经以为长期利率反映未来短期利率的期望值，各国的利率差别反映了汇率贬值的期望值。而今，我们看到在债券和外汇市场以至股票市场中，有随时间改变的风险溢价。我们曾经以为共同基金的平均收益由 CAPM 来很好解释。而今，我们看到，基金可能挣得不被 CAPM 所解释的平均收益，即与市场风险无关的、根据投资“风格”变化的平均收益。”作者接着综述了这些年来的各种新实证研究成果，并且正确地指出，“这些观点不是意识形态或者学术教条的信念。更多的是它们综述了小心翼翼的经验研究在四分之一世纪中的发现。”在 Malkiel [25] 和 Shiller [30] 中，除列举类似的研究成果以外，他们更多地谈到“非理性定价”和“非理性投资者”。尤其是 Shiller [30] 以很大的篇幅来阐述行为金融学的兴起。其中当然少不了我们已经在前一讲中提到过的心理学家 Kahneman 和 Tversky 的研究进入金融学。 Shiller 在他的结语中说：“我们不应该期

望市场有效性会过分失常，使得直接利润源源不断地滚滚而来。但是市场有效性在某种意义下可能过分失常。例如，有效市场理论会导致对诸如重大股市泡沫那样的事件的极为不正确的解释。”Shiller 最后还反驳了 Fama [11] 对行为金融学的批评。粗糙地说， Fama 似乎希望行为金融学研究发现的反常只不过反映了一些统计上的波动。而 Shiller 认为：“反常有时随时间而消失或变号的少量事实并非是市场完全理性的证据。”“在进一步的研究中，重要的是要牢记已经展现的有效市场理论的弱点以及主张一种折衷的方法。当有效市场的理论模型已经作为一个理想的世界的解释或特征有其地位时，我们不可能再以其纯粹的形式把它作为实际市场的精确描述来维持。”Shiller 的这番话应该认为是比较中肯的。

9.4 对于我们的讲义所讨论的范围来说，我们更感兴趣的是理论上怎样为有效市场假设建立模型。弱有效性假设基本上与无套利假设是一回事。这是我们始终在讨论的问题。尤其是我们的讲义以无套利假设、线性定价法则和资产定价基本定理为中心。金融市场的一般经济均衡框架，也已经为无套利假设建立了理论根据。在 Magill–Quinzii [24] 中，还把模型进一步动态化，把二期模型扩充为多期模型。由此可以把弱有效性的一般均衡理论表达得更为清楚。这里我们对此不再作深入介绍，但是我们可以想象这一理论大致是怎样的结构。有兴趣的读者可以直接去参看 [24]。

我们更感兴趣的是怎样为半强有效和强有效假设建立一般均衡模型。这样的模型是存在的，它就是著名的**理性预期均衡** (Rational Expectation Equilibrium, REE)。这里首要的问题是怎样来定义投资者的信息集。

在金融经济学中，“信息”是与“风险”同样容易引起误解的概念。“风险”的字面意思是“可能有的危险”。在金融市场中，风险又被细化为市场风险、信用风险、流动性风险、运营风险等。市场风险又被细化为股市风险、债市风险、利率风险、汇率风险等。总之，一切可能引起危险的场合都隐含着风险，以至人们一看到“风险”两字就会浮想联翩，以为它能把各种场合的风险内涵都能包括在内。然而，在金融经济学中，风险被简单地概括为不确定性，而有关风险的研究也被简单地处理为不确定性的一种度量。例如，在 Markowitz 理论中，风险被定义为证券组合的收益率的方差，在资本资产定价模型 (CAPM) 中，风险可以用 β 来刻划，在决策问题中，Arrow–Pratt 风险厌恶度量则是期望效用函数对不确定性的姿态反映等等。这些有关风险的概念与上述种种风险既有一定联系，又不能随意往上套。具体的金融风险问题必须具体地与这些理论上的风险概念相联系。

“信息”的字面意思比“风险”要多得多。信息通常意味着知识、消息、数据等。于是信息集会被理解为辞书、报刊、通讯、数据库等等。信息又经常强调其更新、揭密的特点。于是信息集又会被人们理解为股市行情、公司年报、内部通报之类。信息还会被理解为使人改变看法和信念的信号 (尤其是传闻、谣言之类)，于是信息集又会变成一种控制行为的信号集合。信息还可能是各种可能性的描述。例如，面临一个棋局，那么有关某个棋手的信息通常意味着这一棋手可能下出的各种棋。这时，信息集又可能是棋手的策略集。此外，信息还有公开的、私下的、一阶的、高阶的等等之分。我们在第一讲的附录中讲的寓言，特别体现了同样的信息在不同场合下的不同作用。于是这时的信息集又会形成“共同知识”这样的概念。用这里提到的各种信息集来理解有效市场定义中的信息集都可以作出各种各样的带有“文学色彩”的解释。然而，我们在这里要强调的是：*所有这些解释几乎都不是理性预期均衡意义下的信息集的含义*。

在信息论、信息经济学等学科以及我们这里要谈到的“**信息**”，观念上很像金融经济学中的“风险”，它主要**是对不确定性的一种刻划**。这样的信息观念也类似于金融经济学中的风险观念，既与通常理解的信息含义有一定联系，又有其自己的特定定义。在这样的理解下，我们将会认为：越不确定的“事件”含有的“信息量”越大。这样的观念在上述各种对信息的理解上都不明显。或许，“策略集”的理解稍接近些：我们对某棋手可能出的棋了解得越多，是不是意味着有关他的信息量越大？

现在我们来给出信息集的具体定义。信息集的概念是基于状态集的一个概念。假定我们考虑的问题基于一个可能发生的各种状态构成的集合 Ω. 为了简单起见，我们只考虑 Ω 是有限集的情形。那么所谓“事件”(“某状态发生”、“某几个状态同时发生”、“某几个状态之一发生”、“某状态不发生”等等) 就是这个状态集的子集。在一定的规定下，所有可能发生的“事件”的集合构成 Ω 的上的一个 σ-域。所谓 Ω 的 σ-域是指 Ω 的子集所构成的集合，它对子集的并、交运算和余集运算封闭 (即“事件 A 和事件 B 都发生”、“或者事件 A 发生，或者事件 B 发生”、“事件 A 不发生”等都是事件)。对于元素足够多的状态集 Ω 可能有许多不同的 σ-域。最粗的 σ-域只有两个元素，即只包含空集和全集的 $\mathcal{F}_0 = \{\emptyset, \Omega\}$. 它意味着只存在两个事件：“至少有一个状态发生”和“什么状态也不发生”。最细的 σ-域是 Ω 的所有子集的并集。这时，任何事件都可能发生。Ω 上的一般的 σ-域则界于二者之间。不难看出，有限集 Ω 的每一个 σ-域 $\tilde{\mathcal{F}}$, 都一定对应 Ω 的一种分划，即，$\Omega = \bigcup_{i=1}^{I} \Omega_i$, 且对于任何 $i \neq j$, $\Omega_i \cap \Omega_j = \emptyset$, 使得 $\tilde{\mathcal{F}}$ 就是由这些 Ω_i 的并和交所生成的。所谓 Ω 上的**信息集**，就是指 Ω 上的一个 σ-域，它紧密联系着其对应的 Ω 上的分划。例如，$\Omega = \{1, 2, 3, 4\}$. 那么 Ω 的一种分划可能是

$\Omega' = \{\{1,2\},\{3,4\}\}$. 而其对应的$\sigma$-域是$\mathcal{F}' = \{\emptyset, \{1,2,3,4\}, \{1,2\}, \{3,4\}\}$.

一般来说，任给Ω上的两个σ-域$\mathcal{F}_1$和$\mathcal{F}_2$, 它们之间不能比粗细。但是如果有$\mathcal{F}_1 \subset \mathcal{F}_2$, 我们就称$\mathcal{F}_2$比$\mathcal{F}_1$细。它可以理解为信息集$\mathcal{F}_2$比信息集$\mathcal{F}_1$“有更多的信息”。另一方面，给定两个$\sigma$-域$\mathcal{F}_1$和$\mathcal{F}_2$，不管它们之间能否比粗细，我们有$\mathcal{F}_1 \cap \mathcal{F}_2$是比它们都粗的最大的$\sigma$-域，它意味着二者的“公共信息”；而$\mathcal{F}_1 \cup \mathcal{F}_2$一般来说不一定是$\sigma$-域，但是我们可以把它扩充为包含它的最小$\sigma$-域，并且不仿仍记为$\mathcal{F}_1 \cup \mathcal{F}_2$, 那么它将意味着二者的“信息之和”。在以下的模型中，每个投资者i都有一个信息集$\mathcal{F}^i$. 那么I个投资者的“公共信息”集就是$\bigcap_{i=1}^{I} \mathcal{F}^i$, 他们占有的“所有私人信息”集就是$\bigcup_{i=1}^{I} \mathcal{F}^i$.

这样的信息集的概念其实我们在前面已经遇到过。在二叉树以至更一般的事件树的概念中，我们曾经提到过σ-域流是一种反映“信息越来越清晰”的过程。尤其是在二叉树中，如果有N个时刻，那么总的状态集Ω就是包含2^N个状态的集合。它的最细的σ-域$\mathcal{F}_N$就是由这个2^N个元素的集合的所有子集所构成。它的最粗的σ-域则是第0个时刻由整个Ω生成的两个元素的σ-域$\mathcal{F}_0$. 而界于二者之间的第k个时刻的σ-域$\mathcal{F}_k$则是把Ω分划为2^k个子集所生成的σ-域$\mathcal{F}_k$. 随着时刻的增大，这些σ-域 (信息集) 就越来越细。

有了信息集的概念以后，我们就可用它来刻画市场上每个投资者所掌握的信息状况。我们再回到我们以前经常讨论的模型上：当前和未来两个时刻，$K+1$种证券，S种未来状态，I个投资者等等。但是我们不再认为当前状态是确定的，而是认为每个投资者i都有包含S个元素的状态集Ω上的σ-域$\mathcal{F}^i$作为它的信息集。证券k的当前价格$p(x_k)$也不再是确定量，而是一个随机变量 (即S维向量)。由于它也能生成Ω的一种分划 (价格一样的状态形成Ω的一个子集)，从而它也能生成它自己的信息集$\mathcal{F}^p$. 这就是“价格带来的信息”。认为当前也是不确定的似乎不很好理解。其实这只是为了模型的简化。更合理的模型至少是三时期：事前、中期、事后，这里中期和事后都是不确定的。把当前看作事前和中期的合成只是一个简化模型。

在这样的前提下，投资者i将考虑第七讲中类似的证券组合最优选择问题：求证券组合θ, 使得

$$\left\{\begin{array}{ll} \max & E^i[u(z^{i0}, z^{i1})] \\ \text{s.t.} & z^{i0} = \omega^{i0} - \sum_{k=0}^{K} \theta_k p(x_k), \\ & z^{i1} = \omega^{i1} + \sum_{k=0}^{K} \theta_k x_k. \end{array}\right. \tag{1}$$

与以前的不同之处在于第二个等式是随机量的等式，第一个等式也是随机量的等式。而E^i是投资者i的“主观概率”下的数学期望。其中对z^{i1}求数学期望与以前没有多大区别，但是概率可以是投资者i所独

有的。但是对于 z^{i0} 的数学期望有点特殊。这里初始持有 ω^{i0} 应该是个 $\mathcal{F}^i$-可测的随机变量，即它对于 $\mathcal{F}^i$ 对应的 Ω 的子集分划来说，在分划的每一子集上它是常数。另一方面，价格 $p(x_k)$ 又是 $\mathcal{F}^p$-可测函数。于是 z^{i0} 应该是个 $\mathcal{F}^i \cup \mathcal{F}^p$-可测函数。投资者 i 对于 z^{i0} 的“主观概率”应该对于这个 σ-域 $\mathcal{F}^i \cup \mathcal{F}^p$ 来定义。[3]

现在假设有 I 个这样的投资者，它们与前面的一些假设一起形成一个金融市场 $\mathcal{E}(u^i, \omega^i, \Omega)$. 类似于以前的讨论，我们可以定义它的一般均衡 $(\bar{\theta}^i, \bar{p})$ 为满足下列两条件的证券组合－价格组：

1. 对于任何 $i = 1, 2, \ldots, I$, $\bar{\theta}^i$ 是当 $p = \bar{p}$ 时的问题 (1) 的解。
2. $\bar{\theta}^1 + \bar{\theta}^2 + \cdots + \bar{\theta}^I = 0$.[4]

这样的均衡就称为**理性预期均衡**。理性预期均衡是否存在当然也是一个需要深入讨论的问题。事实上，通常的金融市场的一般均衡存在问题只是它的一个特例 (所有信息集 $\mathcal{F}^i$ 和 $\mathcal{F}^p$ 都是最粗的信息集)，其中尤其是包括不完全市场的的一般均衡存在问题。因此，一般来说，理性预期均衡是不一定存在的。但是如果均衡存在，那么它的均衡定价函数所提供的信息集，就可用来刻画有效市场。

假定对于某市场的理性预期均衡价格 $\bar{p}$ 存在，这里 $\bar{p}$ 是定义在证券的未来价格上的随机变量值函数。对于一个信息集 $\mathcal{F}' \subset \mathcal{F}$ 来说，**市场 $\mathcal{E}$ 称为对于 $\mathcal{F}'$ 有效**是指：对于任何 $i = 1, 2, \ldots, I$, $\mathcal{F}' \subset \mathcal{F}^i \cup \mathcal{F}^{\bar{p}}$. 这样一来，有效市场的分类就可表达得更加明确。公共信息集可定义为 $\bigcap_{i=1}^I \mathcal{F}^i$. 它必定是市场有效的。这种有效性就是半强有效性。私人信息集全体定义为 $\bigcup_{i=1}^I \mathcal{F}^i$. 因此，强有效可定义为

$$\forall i = 1, 2, \ldots, I, \qquad \bigcup_{i=1}^{I} \mathcal{F}^i \subset \mathcal{F}^i \cup \mathcal{F}^{\bar{p}}.$$

这一条件不一定满足。如果它满足，就意味着这一均衡是一种信息被**完全披露**的理性预期均衡。但是如果所有私人信息都被反映在价格信息中，人们就没有搜集信息的动力，以至就没有任何私人信息。Grossman–Stiglitz 悖论又由此引起。于是在这一框架的意义下，又可以论证，信息被完全披露的理性预期均衡不存在。

---

[3] 如果说这样的说法不够妥当，那么我们可以说，投资者 i 的“主观概率”仅仅是对 $\mathcal{F}^i$ 定义的，而价格则是对“客观概率”而言的。这时，上述对 σ-域 $\mathcal{F}^i \cup \mathcal{F}^p$ 定义的“主观概率”求数学期望，其实是价格先对 $\mathcal{F}^i$ 求条件数学期望，然后再对条件数学期望求 $\mathcal{F}^i$ 的“主观数学期望”。这样求得的数学期望与直接求“主观数学期望”不同，它就是所谓“根据所掌握的信息来改变对可能性的估计”的 Bayes 方法。

[4] 与以前一样，我们也可以把这里的“债券型”证券改成“股票型”证券，即可假设每个投资者都已持有一定的证券组合，而均衡则要求总证券组合不变。

这样，我们就通过理性预期均衡的概念为半强有效与强有效建立理论框架。这里我们还可以看到，在半强有效与强有效之间，还可以有许多中间的市场有效。然而，我们也要注意到，在这一理论框架中的信息概念的特殊性。由此产生的半强有效与强有效与其原来的本意还有很大的区别。

9.5　最后，我们再来看看 CAPM 和 APT 与有效市场的关系。既然 CAPM 和 APT 都有其金融市场的一般经济均衡模型，而通常的一般经济均衡模型又是理性预期均衡模型的特例，那么上述的有效市场的讨论也就适用于 CAPM 和 APT. 甚至我们还可以为它们建立每个投资者都有自己的私人信息集的理性预期均衡模型。但是无论是 CAPM 还是 APT，都有“投资者期望均匀性”的假定，也就是说他们的“主观概率估计”都是一样的，这当且仅当价格完全披露私人信息时才有可能。这样一来，如果说 CAPM 和 APT 都是建立在理性预期均衡的基础上，那么它必须与市场的强有效联系在一起。也就是说，如果市场强有效性不成立，那么 CAPM 和 APT 都不能成立。而反过来却不一定，即，有可能市场强有效性成立，而 CAPM 和 APT 不成立。这样的观点有可能对近年来出现的一些关于市场有效性检验的研究给出新解释，即它们实质上只是否定了一定意义下的 CAPM 和 APT, 而不是有效市场。

注意到这些方面并不是要否定“经典金融经济学”的理论价值，而是要使人们认识到，金融经济学研究还有许多方面需要深入。在经典的理论框架中，我们已经得到资产定价基本定理、Markowitz 理论、CAPM、APT 等等。但是如果再进一步考虑投资者的信息、信念及其引起的行为以及金融市场的微观结构，就会发现经典理论的许多不足。这正是当前金融经济学理论研究的许多新课题。

说明　有效市场假设是金融经济学理论中的一个大题目。如果说金融经济学研究的中心问题是金融商品的定价问题，那么有效市场假设必须是理论研究的前提，否则那些有套利机会的金融商品是不可能为其定价的。当然，还有更精细的研究认为一般经济均衡与市场无效是不矛盾的，那是指把市场看成一个开放系统，由于金融外部条件引起的市场无效。关于有效市场假设是否成立始终在金融界的“实务派”与“学院派”之间争论。近年来，人们发现，无论是要肯定还是否定有效市场假设都不那么容易。尤其是要把检验看作均衡模型和市场有效的联合检验，历史上已有的检验研究很少能通过这样的要求。目前对有效市场的研究已经越来越深入。尤其是 Lo–MacKinley [22], Shleifer [29] 等新书都是以与有效市场理论唱对台戏的姿态出现。这可能给人一种有效市场理论已经过时的错觉。然而，就像相对论力学并未彻底否定经典力学一样，“经典有效市场理论”仍然有其应有的价值。

关于有效市场假设的研究大部分是实证研究，而理论金融经济学是把有效市场假设看作前提的。因此，在不太强调一般经济均衡基础的理论金融经济学著作 Duffie [4], [5], Huang–Linzenberger [13] 和 Ingersoll [14] 中都不讨论有效市场问题。Jarrow

[15] 因为比较强调一般经济均衡基础，所以其最后一章专门讨论了有效市场。我们这里的有关理性预期均衡的叙述基本上取材于它。但是理性预期均衡本身就是个非常重要的概念。我们知道 R. E. Lucas (1937–) 获得 1995 年诺贝尔经济学奖正是因为他发展和应用理性预期假设，从而改造了宏观经济分析以及加深了人们对经济政策的理解。但是在金融经济学中讨论理性预期均衡更多的是 Radner 等人的工作。这一概念正是 Radner 提出的金融市场一般经济均衡模型的推广。Radner [26] 是这方面研究的系统叙述。而 Jordan–Radner [16] 是理性预期均衡存在条件研究的一个综述。不过 Lucas [23] 也曾为有效市场提出过一个简单的理性预期均衡模型。LeRoy [18] 则是一篇更多地从理论上来理解有效市场的综述论文，值得一读。

用理性预期均衡来讨论金融市场中的信息传递常使人有不自然的感觉，这是因为信息集用 σ-域来定义似乎与通常理解的信息集有较大的距离。另一种类似的讨论是把投资者的主观概率看作价格的函数，由此来讨论通过价格转递信息的问题。这样可得到所谓 Green–Lucas 均衡的概念。有关讨论可参看 Laffont [17]，该书也是信息经济学方面的名著。

关于有效市场理论的研究讨论目前很多。尤其是伴随着行为金融学的兴起，这方面的争论更是十分热闹。我们在前面已经介绍了不少有关的论述。附录中，我们还专门翻译了一段 Lo–MacKinlay 的论述，把它作为一种目前比较典型的观点向大家推荐。至于行为金融学与有效市场理论的关系论述，读者可参看 Shleifer [29] 或 Shiller [30]。

思考题和习题

1. 什么是有效市场假设 (EMH)？你对此有何看法？
2. 世界各国的证券市场都流行基本分析、技术分析方法？你认为这类方法是否有其存在价值？有人认为，这类方法适用于短线运作，而有效市场假设适合长线运作，对此，你有什么看法？
3. 什么是随机游走假设？它与弱形式的有效市场假设有什么关系？
4. 什么是 σ-域？怎样来理解用它来定义投资者的信息集？
5. 什么是理性预期均衡？怎样用它来定义半强形式和强形式的有效市场？
6. CAPM 和 APT 与有效市场假设有何关系？

参考文献

[1] Campbell, J. Y., A. W. Lo & A. C. MacKinlay, 1997, The Econometrics of Financial Markets, Princeton University Press, Princeton, New Jersey.

[**2**] Cochrane, J. H., 1999, New facts in finance, *Economic Perspectives, Federal Reserve Bank of Chicago*, **23**: 36–58.

[**3**] Cowles, A., 1933, Can stock market forecasters forecast? *Econometrica*, **1**: 309–324.

[**4**] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[**5**] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N. J..

[**6**] Fama, E., 1965, Random walk in stock market prices, paper number 16 in the series of *Selected Papers of the Graduate School of Busines, University of Chicago,* reprinted in the *Financial Analysts Journal* (September-October 1965), *The Analysts Journal*, London (1966), *The Institutional Investor,* 1968.

[**7**] Fama, E., 1965, The bdhavior of stock market prices, *Journal of Business*, **38**: 34–105.

[**8**] Fama, E., 1970, Efficient capital markets: a review of theory and empirical work, *Journal of Finance*, **25**: 383–417.

[**9**] Fama, E., 1976, Efficient capital markets: reply, *Journal of Finance*, **31**: 143–145.

[**10**] Fama, E., 1991, Efficient markets: II, *Journal of Finance*, **46**: 1575–1617.

[**11**] Fama, E., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics*, September, **49:3**, 283–306.

[**12**] Grossman, S. J. & J. E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review*, **70**: 393–408.

[**13**] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[**14**] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[**15**] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[16] Jordan, J. and R. Radner, 1982, Rational expectations in microeconomic models: an overview, *Journal of Economic Theory*, **26**: 201–223.

[17] Laffont, J.-J., 1988, Economics of Uncertainty and Information, MIT Press, Cambridge, Massachusetts.

[18] LeRoy, S. F., 1973, Risk aversion and matingale property of stock prices, *International Economic Review*, **14**: 436–446.

[19] LeRoy, S. F., 1976, Efficient capital markets: comment, *Journal of Finance*, **3**: 139–41.

[20] LeRoy, S. F., 1989, Efficient capital market and martingale, *Journal of Economic Literature,* **27**: 1583–1621.

[21] Lo, A. W. & A. C. MacKinlay, 1988, Stock market prices do not follow random walks: evidence from a simple specification test, *Review of Financial Study*, **1**: 41–66.

[22] Lo, A. W. & A. C. MacKinlay, 1999, A Non-Random Walk Down Wall Street, Princeton University Press, N. J..

[23] Lucas, R. E. Jr., 1978, Asset prices in an exchange economy, *Econometrica*, **46**: 1429–1445.

[24] Magill, M. & M. Quinzii, 1997, Theory of Incomplete Markets, Vol. 1, MIT Press, Cambridge, Massachusetts.

[25] Malkiel, B. G., The efficient market hypothesis and its critics, *Journal of Economic Perspectives*, **17:1**, 59–82.

[26] Radner, R., 1982, Equilibrium under uncertainty, in K. J. Arrow and M. D. Intrilligator eds., *Handbook of Mathematical Economics*, Vol. II. Amesterdam: North–Holland, 923–1006.

[27] Ross, S., 1987, The interrelation of finance and economics: theoretical perspectives, *Amer. Econ. Rev.,* **77**: 29–34.

[28] Samuelson, P., 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review*, **6**: 41–49.

[29] Shleifer, A., 2000, Inefficient Markets, An Introduction to Behavioral Finance, Oxford University Press, Oxford.

[30] Shiller, R. J., 2003, From efficient market theory to behavioral finance, *Journal of Economic Perspectives*, **17:1**, 83–104.

附录 (摘译自 [22], 6–8 页)：**有效市场假设的现状**

有一个在经济学家们中广泛流传的老笑话，说是一位经济学家与他的同伴在马路上漫步，发现地上有一张 100 美元。正当他的同伴要伸手把它捡起来时，经济学家说：“何必为此操心？如果它真是 100 美元，早就被人捡走了。”

这个经济学逻辑的幽默例子，对于有效市场假设 (Efficient Markets Hypothesis) 的研究者来说，可说是歪打正着，击中要害。有效市场假设是在整个社会科学中最被人争论不休、反复研究的命题之一。它的陈述出奇地简单，对于学术探索和商业实践都有意味深长的推论，而对反驳它的经验证明仍然有令人惊讶的反弹。即使经过三十年的研究以及数以千计的论文发表以后，经济学家们仍然对于市场 (特别是金融市场) 是否有效没有达成共识。

我们对于有效市场假设能够得到什么结论呢？叫人困惑的是，在金融经济学家之间也还没有共识。尽管在围绕有效市场假设的统计分析、数据库以及理论模型上有许多进展，在这场论战中的已有的大多数经验研究的主要影响在于从各方面来坚定倡导者们的决心。

这一事态的缘由之一是有效市场假设自身并非是一个适当定义和经验上可被反驳的假设。为使它可操作，人们必须规定附加结构，例如投资者偏好，信息结构，商业条件等等。然后，一个有效市场假设的检验就变为多个辅助假设的检验，而对这样一个联合假设的检验的拒绝很少告诉我们是联合假设的哪一方面与数据不相容。股票价格过于波动是否是因为市场无效，还是由于风险厌恶，或者分红调整？所有三个推断都与数据相容。尤其是，为区分它们而设计的新的统计检验无疑又要求它们自身的辅助假设，而它们又可能转而成为问题。

最为重要的是，有效市场假设的检验可能不是衡量给定市场的效率的最说明问题的手段。更说明问题的是一个指定的市场相对于另一些市场的**相对**有效性，例如期货市场相对于现货市场，拍卖市场相对于零售市场等等。相对有效性概念与全有全无的绝对有效性相比的好处在于容易通过类比的方式来辨认。物理系统经常基于能量或燃料转化为有用功的相对比例来给出其效率级别。因此，一个活塞发动机可被评级为 60% 的效率，它意味着包含在发动机燃料中大致有 60% 的能量用来转动机轴，其余的 40% 则以其他的功的形式，例如热、光、噪声等等变为损失。

很少有工程师考虑进行统计检验，以决定发动机是否完全有效；

那种完全有效的发动机只存在在理想的无摩擦的想象世界中。但是测量相对效率 (相对于无摩擦的理想物) 则是常识。事实上，我们对于许多家用产品：空调、热水器、冰箱等等，我们都会期待这样的测量。因此，从实际观点来看以及根据 Grossman and Stiglitz (1980)，有效市场假设是一个理想化，它在经济上是不可实现的，但是它可用来作为测量相对效率的有用的基准。

有效市场假设的更实际的文本是用某种类比来提出的，它需要统计力学中的热平衡的概念。尽管偶尔有“超额”利润的机会，但就平均和长时间而言，没有某种类型的竞争优势，例如，专有信息、专有技术、金融创新等等，不可能持久地赢得这样的利润。换句话说，在一个有效市场中，仅有的赢得**持久**的正利润的办法就是发展竞争优势；在这种情形下，利润可以看作产生这种竞争优势的经济寻租。这种利润的持久性是一种重要的资质——在有效市场的这一文本中，一顿偶然的免费午餐是允许的，但是免费午餐计划是被排除的。

为看出有效市场假设的这样的一种解释为什么更为实际，暂且先来考虑对例如生物技术那样的非金融市场运用有效市场假设的经典文本。比如，考虑的目标是开发用于艾滋病的疫苗。如果生物技术市场在经典意义下是有效的，那么这样的疫苗不可能被开发——如果可能，早就有人开发过了！这显然是一个颟顸可笑的假定，因为它无视在生物技术的研究开发中的困难和孕育延迟。尤其是，如果一家药厂成功地开发出这样的疫苗，其赢得的利润将会以几十亿美元来计。难道这能被当做“超额”利润或者生物技术专利所产生的经济寻租？

金融市场在原则上并无区别，而仅仅在程度上有所区别。因此，专业投资所产生的利润并不需要市场是**无效的**，并且可能干脆就是对金融技术上的突破的公平的奖赏。不管怎么说，很少有分析家把 Amgen 公司过去几年的丰厚利润看作药品市场无效的证据——Amgen 近年来的利润容易通过几种新药 (例如， Epogen, 一种促进红血细胞产生的药品) 的开发来辨别，它们被看作生物技术的突破。与此类似，即使在有效的金融市场中，对于金融技术的突破，同样会有堂而皇之的收益。

当然，由于进入市场的门槛通常很低，竞争的程度却非常激烈，以及大多数的金融技术没有专利 (虽然这可能很快就会改变)，因而金融创新的可获利性的“半衰期”相当短。这些特点蕴含着金融市场应该是相对来说更为有效，并且事实上也确实如此。“二手证券”市场比起二手车市场来要有效得多。但是断言金融市场必须完全有效，就像宣称艾滋病疫苗不可能发现一样。在一个有效市场中，很难赚大钱，但并非不可能。

十、连续时间金融学

本讲要求：对连续时间金融学的基本数学框架有所了解。对随机分析的基本概念：随机游走、Brown 运动、鞅、随机积分、随机微分方程等有所了解。由此再了解连续时间的 Black–Scholes 期权定价模型以及等价概率鞅测度、 Black–Scholes 方程等概念。再进一步了解利率期限结构的连续时间模型。

数学预备知识：首先要对概率论的基本概念相当熟悉。然后，再通过基本概率论的概念来理解随机游走、 Brown 运动、随机积分、Itô 公式等概念。彻底搞清这些数学理论不是本讲义的任务，但是希望通过本章的介绍，能理解这些数学概念的含义及其所反映的金融学内容。也要对微分方程理论有初步的概念，由此理解 Black–Scholes 方程意味着什么，其面临的数学问题是什么。

内容：

- 10.1 Brown 运动、随机分析等的一些启发性叙述。
- 10.2 随机分析的进一步叙述。
- 10.3 连续时间的 Black–Scholes 模型和期权定价公式。
- 10.4 Black–Scholes 公式原来的推导。
- 10.5 利率期限结构的连续时间模型。

“连续时间金融学 (continuous-time finance)”是 Merton 在 1990 年初版、1992 年修正再版的重要著作 ([15])。不同于我们前面讨论的二时期模型或多时期模型，连续时间金融学的模型中都假定时间是连续变化的。这一方面便于许多强有力的数学工具 (例如微分方程、随机分析等) 的应用，从中立即可以得到许多深刻而又简洁的结论；另一方面，当前的金融市场瞬息万变，证券交易的速度越来越快，以任何时间间隔来离散化的数学模型都不足以反映这种变化，以至连续模型比离散模型更为接近实际。因此，今天如果对连续时间金融学一无所知，就很难被认为对金融经济学有全面的了解。这里涉及的数学工具虽然比较艰深，但是对它们所反映的基本思想还是不难理解的。从经济思想来看，连续模型与离散模型并无多大区别。

连续时间金融学自然又要追溯到 1900 年 Bachelier 的学位论文。就是他首先在对证券市场的研究中，把时间连续变化的随机过程与股票价格的变化联系在一起。同时，正如我们在前几讲中所提到的，Bachelier 提出了 Brown 运动的数学定义。这一讲，我们就先从 Brown 运动的不太严格的数学定义讲起。

10.1　Brown 运动是随机游走概念的连续化。因此， Brown 运动是随机游走的极限情况。随机游走是一个随机序列，即它是一系列随机

变量的排列。一个标准随机游走的起点项是 0，第 1 项是一个或取 +1, 或取 −1 的随机变量，并且取 +1 或 −1 的概率都是 1/2. 不妨称这样的随机变量为“随机摆动”。第 2 项是从第 1 项的位置出发的“随机摆动”，因此，它是“随机摆动”的两次独立试验之和。同理，第 3 项是“随机摆动”的 3 次独立试验之和。一般情形下，设 $(W_n)_{n=0,1,2,\dots,}$ 是标准随机游走序列，那么

$$W_0 = 0, \qquad W_{n+1} = W_n + \varepsilon_n, \qquad n = 1, 2, \dots,$$

其中 ε_n, $n = 1, 2, \dots,$ 是相互独立的以概率 1/2 分别取 ±1 的“随机摆动”。很明显，随机游走序列有以下这样的性质：

1. $E[W_n] = 0$, $n = 0, 1, 2, \dots$, $\mathrm{Var}[W_n] = n$.
2. 如果 $m > n \geqslant 0$, 那么 $W_m - W_n$ 与 W_k, $k = 1, 2, \dots, n$, 相互独立。
3. 如果 $m > n$, 那么 $W_m - W_n$ 与 $W_{m-n} - W_0$ 有同样的概率分布。

作为随机游走序列的连续化，标准 Brown 运动 $(B_t)_{t\geqslant 0}$ 是一个随机过程，即它是随时间 t 变化的随机变量，并且有下列性质：

1. (**连续性**)：(B_t) 的几乎所有的轨线是连续的；
2. (**增量的独立性**)：如果 $s < t$, 那么 $B_t - B_s$ 对所有 $u \leqslant s$ 的 B_u 独立；
3. (**增量的平稳性**)：如果 $s < t$, 那么 $B_t - B_s$ 与 $B_{t-s} - B_0$ 有同样的概率分布；
4. (**正态性**)：对于任何 $t \geqslant 0$, B_t 是满足 $E[B_t] = 0$ 和 $\mathrm{Var}[B_t] = t$ 的服从正态分布的随机变量。

这里其实只有前三条是 Brown 运动的定义，其中第一条是必要的数学假设，以排除各种“跳跃”运动。第四条中的正态性可以由前三条导出，就像由引理 8.1 (中心极限定理) 可指出，作为若干独立的“随机摆动”叠加的随机游走序列，在某种意义下以服从正态分布的随机变量为极限。方差与时间成正比也可导出。余下的则是标准化假设。

另一方面， Brown 运动 (B_t) 的增量过程 $(V_t^s)_{t\geqslant 0} = (B_{t+s} - B_t)_{t\geqslant 0}$ 满足 $E[V_t^s] = 0$,

$$E[V_{t+\tau}^s V_t^s] = \begin{cases} 0, & \tau \geqslant s, \\ R(\tau), & \tau < s, \end{cases}$$

这里 $R(\tau)$ 是 τ 的函数，尤其是 $R(0) = \mathrm{Var}[B_s] = s$. 如果一个随机过程 f_t 满足： $E[f_t] = \mu$ 以及 $E[f_{t+\tau} f_t] = r(\tau)$ 都与 t 无关，那么它称为 (**广义**) **平稳过程**。Brown 运动的增量过程就是一个平稳过程。平稳过程可以粗糙地说是一种统计特性不随时间变更的随机过程。这种随机过程的最重要的特性是它一定是许多“随机振荡”与一个常数的叠加。所谓“随机振荡”就是形式为 $g_t = a \sin \omega t + b \cos \omega t$ 的随机过程，其中 ω (频率) 是常数，a, b 是数学期望为零的随机变量，而 a 和 b 的方差之和则代表这一“随机振荡”的强度。一个平稳随机过程一定是有限个或者无限个 (求和或者求积分) 随机振荡的叠加。如果它的“随机振荡成分”中有一个“成分”的“强度”特别强，那么这个平稳过程就显示出一定的周期性。这种方法有时也被用来分析股价 (或股票收益) 的周期性，但是这样做以前，首先要判别股价序列 (或股票收益序列) 是否是平稳的，或者要设法使其“平稳化”(例如去掉其中的增长趋势)。Brown 运动的增量过程是一个“频谱”很广的平稳过程，即它的所有“随机振荡成分”的“强度”都差不多。所有频率的随机振荡成分的强度都一样的平稳过程称为“白噪声”(称它是“白”的是借用“白色光”为包含各种颜色 (频率) 的光的含义) 。 Brown 运动的增量过程很接近于“白噪声”。一种数学上不太严格的说法认为，Brown 运动对时间的“导数”是“白噪声”。

与随机游走序列一样，Brown 运动也可以与鞅的概念联系在一起。这里自然也需要 σ-域流、适应过程之类的概念。其细节我们将在下面给出。这里我们暂且先粗糙地来理解它。首先，如同随机游走序列一样， Brown 运动是一个鞅 (随机过程)，即对于 $s < t$ 来说，B_t 关于时刻 s 的 σ-域 $\mathcal{F}_s$ 的条件期望就是 B_s. 或者用普通语言说，在时刻 s 对未来时刻 t 的 Brown 运动来估计，恰好就是时刻 s 的 Brown 运动。

由标准 Brown 运动还可生成其他的鞅。最简单的情形是 $(aB_t + b)$, 这里 a, b 为常数。它显然也是鞅。稍复杂的情形是进行平方运算。但是 (B_t^2) 不是鞅，因为 $B_0^2 = 0$, 但 $E[B_t^2] = t \neq 0$. 然而，可以验证，$(B_t^2 - t)$ 是鞅。这种鞅称为**平方鞅**。另一种运算是指数运算。对 Brown 运动简单的指数运算也不能生成鞅。例如， $(\exp B_t)$ 就不是鞅。然而，$(\exp(\sigma B_t - (\sigma^2/2)t))$ 是鞅，这里 σ 是常数。这种鞅称为**指数鞅**。对它们的论证将在下面给出。

在金融经济学中，经常简单地假定股票价格 S_t 的变化满足下列随机微分方程：

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_t, \tag{1}$$

这里 μ 称为**漂移** (drift)，σ 称为**波动率** (volatility)。这一方程的严格数学含义有待在后面明确，但是我们在这里可以粗糙地把它理解为：左端是随机变量证券的“瞬时收益率”，它的均值是 μdt, 它的方差是 $\sigma^2 dt$

(因为按不太严格的说法，$dB_t = B_{t+dt} - B_t$ 的方差是 dt)。也就是说，这里沿袭了 Markowitz 用均值—方差来刻划收益率的传统做法。同时，由于 Brown 运动的特性，“瞬时收益率”还是一个正态分布的随机变量。另一方面，把上述方程在形式上写成

$$\frac{dS_t}{dt} = S_t\left(\mu + \sigma\frac{dB_t}{dt}\right),$$

我们还可以说，股价的增长率 μ 受到强度为 σ 的“白噪声”的干扰。这样定义的 (S_t) 也可直接表示为 $(\exp(\sigma B_t + (\mu - \sigma^2/2)t))$. 一般来说，它不是鞅。但是如果 $\mu = 0$, (S_t) 就变成指数鞅 $(\exp(\sigma B_t - (\sigma^2/2)t))$. 由此还可以看出，对于任何 t 来说，随机变量 S_t 服从对数正态分布。

对股票价格作这样的假设并不是凭空而来的，而是许多金融计量经济学家们长期分析数据所得到的结果。我们把上述随机微分方程离散化 (这可不太严格地认为是令“$dt = 1$”而得到的)，则有

$$\frac{S_{n+1} - S_n}{S_n} = \mu + \sigma\varepsilon_{n+1}, \tag{2}$$

这里 ε_n 是随机摆动序列。这就是说，方程 (1) 实际上出自“**股票价格的收益率序列 (去除固定收益后) 形成一个随机摆动序列**”这一假设。历史上，曾经有人认为市场弱有效性假设等价于股票价格序列形成一个随机游走序列，其首创者自然也是 Bachelier。但是这种说法已经在理论和实证两方面都被否定。实际上，大量实证分析指出，股价差分序列通常不是平稳序列，而股价收益序列才是平稳序列，因而才可能形成随机摆动序列。另一方面，(1) 还可记作

$$d\log S_t = \mu dt + \sigma dB_t. \tag{3}$$

它的离散化是

$$\log S_{n+1} - \log S_n = \mu + \sigma\varepsilon_{n+1}. \tag{4}$$

它意味着**股票价格的对数差分序列形成一个随机摆动序列**。[1] (3) 与 (4) 并不一致，但是它们在“一阶近似”的意义下是一致的，即我们有近似等式：

$$\log\left(\frac{S_{n+1}}{S_n}\right) \approx \frac{S_{n+1} - S_n}{S_n}.$$

这一近似等式来自：当 $|x|$ 很小时，$\log(1+x) \approx x$. 因此，当 S_{n+1} 很接近 S_n 时，我们就可把 $\log\left(\frac{S_{n+1}}{S_n}\right)$ 当作股票收益的近似值。实际上，由

---

[1] 然而，1988 年 Lo 和 MacKinlay [13]又用更精确的实证分析否定了这一结果。也就是说，经典的 Black–Scholes 公式中的有关假设精确地说是不成立的。

于在数学上，$\log\left(\frac{S_{n+1}}{S_n}\right)$ 比较容易处理，在实证分析时，经常就把 (净) 收益率这样定义。

现在我们来考虑一般的随机微分方程：

$$dX_t = K_t dt + H_t dB_t. \tag{5}$$

这里 (K_t) 和 (H_t) 是两个随机过程 (在上述的股价方程中，$K_t = S_t\mu$, $H_t = S_t\sigma$)；(X_t) 则是由此定义的随机过程。我们现在要考虑，对于 X_t 的函数 $f(X_t)$ 应该满足怎样的随机微分方程。如果上述方程中 $H_t = 0$, K_t 是 t 的普通函数，那么这就是复合函数求导数的问题，即我们有

$$df(X_t) = f'(X_t)dX_t = f'(X_t)K_t dt. \tag{6}$$

但是在 H_t 不等于零时，情况不再如此。

要说原因还要从微分的定义说起。微分表达式 $df(x) = f'(x)dx$ 的实际含义在于：当 x 的增量 Δx 变化很小时，$f(x)$ 的增量 $\Delta f(x) = f(x+\Delta x) - f(x)$ 可表示为 $f'(x)\Delta x + o(\Delta x)$, 其中 $o(\Delta x)$ 表示 Δx 的高阶无限小，即 $\lim_{\Delta x\to 0} o(\Delta x)/\Delta x = 0$. $f'(x)\Delta x$ 作为 $\Delta f(x)$ 的“线性主部”称为 f 在点 x 处的微分 $df(x)$. 再由于 dx 就等于 Δx, 就得到上述的微分表达式。如果 x 还是 t 的函数 $x = x(t)$, 那么我们还有：$df(x(t)) = f'(x(t))x'(t)dt$. 这是因为 $\Delta f(x(t))$ 对于 Δt 的“线性主部”可以由它关于 Δx 的“线性主部”和 Δx 关于 Δt 的线性主部来合成。

然而，对于随机微分方程来说，情况不再如此。事实上，例如，在 K_t 和 H_t 都是 t 的定常函数的情形下，按不太严格的意义，近似等式 $\Delta X_t \approx K_t\Delta t + H_t\Delta B_t$ 的右端意味着一个均值为 $K_t\Delta t$, 标准差为 $H_t\sqrt{\Delta t}$ 的正态分布的随机变量。于是随机变量 ΔX_t^2 的均值中还包含 Δt 的项。这样，(6) 不再成立。对于光滑函数 $f = f(x)$, 利用它的 Taylor 展开式：

$$\Delta f(x) = f(x+\Delta x) - f(x) = f'(x)\Delta x + \frac{f''(x)}{2}\Delta x^2 + o(\Delta x^2),$$

我们就可导得下列 Itô (伊藤) 公式：

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)H_t^2 dt.$$

这也就是说，$df(X_t)$ 作为随机变量，它的均值比 $f'(X_t)dX_t$ 的均值还要多出一个包含 f 的二阶导数的项。这是随机微分学与普通微分学的根本不同：复合函数的微分公式中包含二阶导数项。

我们用 Itô 公式来计算一些简单的函数：

1. $f(x)=x^2$, $dX_t=dB_t$, $X_0=0$. 那么 $X_t=B_t$,

$$df(X_t)=2B_tdB_t+\frac{1}{2}\cdot 2dt.$$

这就是说，

$$B_t^2-t=2\int_0^t B_sdB_s.$$

可以指出，右端是个鞅。很明显，这个公式与非随机情形的区别在于多了一项 $-t$.

2. $f(x)=\log x$, $dS_t=S_t(\mu dt+\sigma dB_t)$, $S_0=x_0$. 于是

$$S_t=x_0+\int_0^t S_s(\mu ds+\sigma dB_s),$$

利用 Itô 公式，可得

$$\log(S_t)=\log(x_0)+\int_0^t\frac{dS_s}{S_s}-\frac{1}{2}\int_0^t\frac{1}{S_s^2}\sigma^2S_s^2ds,$$

以至

$$\log(S_t)=\log(x_0)+\left(\mu t-\frac{\sigma^2}{2}t+\sigma B_t\right),$$

或

$$S_t=x_0\exp\left(\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma B_t\right).$$

这是我们上面已经提到过的结果。

以上是一些以不太严格的形式所叙述的数学结果。如果对这些结果有足够的理解，那么对一般的连续时间金融学的数学表达就会感到并不太难理解。

10.2 下面我们先把上面提到的一些数学概念用较严格的数学形式来表达，然后再来讨论 Black–Scholes 模型。

我们从一个可测空间 $(\Omega,\mathcal{F})$ 出发，这里 Ω 是一个任意的集合，$\mathcal{F}$ 是由 Ω 中的子集所构成的 σ 域。与离散情形类似，为了反映信息量随着时间增长的不断增长，引入 σ-域流 (filtration): $(\mathcal{F}_t)_{t\geqslant 0}$, 这里每个 $\mathcal{F}_t$ 都是 Ω 的子集所构成的 σ-域，并且满足单调性：

$$0<t_1<t_2\implies\mathcal{F}_0\subset\mathcal{F}_1\subset\mathcal{F}_2\subset\mathcal{F}.$$

这是我们在前一讲讨论中所提到的由“事件树”所形成的 σ-域流的推广。但是我们仍能象理解“事件树”那样来理解它。如果在可测空间 $(\Omega, \mathcal{F})$ 中还定义了概率测度 $\mathbf{P}$, 那么同样如前一讲最后讨论的那样，通常还假定零概率事件属于每个 $\mathcal{F}_t$，以避免出现一些实际中很少发生的怪现象。于是代表当前信息的 $\mathcal{F}_0$ 就不仅是一些完全肯定的事件 ($\mathcal{F}_0$ 只包含 Ω 和 $\emptyset$), 还包含零概率事件和概率为一的事件。不过与我们以前的讨论一样，我们以后关心的是如何来确定一些有特殊意义的等价概率测度，从而在我们的框架中，重要的是要明确，零概率事件全体是一个确定的、所有 $\mathcal{F}_t$ 的子集。

有了可测空间 $(\Omega, \mathcal{F})$，就可定义随机变量 X (即 X 为定义在 Ω 上的可测函数，它对于任何实数 c, $\{\omega \in \Omega \,|\, X(\omega) < a\} \subset \mathcal{F}$, 或者说 $\{X < c\}$ 是一个事件) 和随机过程 $(X_t)_{t \geqslant 0}$ (即对于每个 t, X_t 是随机变量)。一个随机过程 (X_t) 对于固定的 $\omega \in \Omega$ 来说, $t \mapsto X_t(\omega)$ 是 t 的函数，它称为该过程的**轨线**。轨线自然不一定是连续的。但是为了减少数学上的非本质的困难，我们以后总假定所讨论的随机过程的几乎所有的轨线都是连续的。这对于证券市场的初步讨论也已足够。

与前面一样，我们同样可以定义适应过程。随机过程 $(X_t)_{t \geqslant 0}$ 称为**适应过程**是指对于任何 t, X_t 是 $\mathcal{F}_t$-可测的。适应过程都是针对 σ-域流而言的。但是对一个固定的随机过程来说，由它也可生成一个 σ-域流，即令 $\sigma(X_s, s \leqslant t) = \mathcal{F}_t$ (它表示对于 X_s $(s \leqslant t)$ 来说可能发生的事件集合所生成的 σ-域)，那么 (X_t) 对于这个 σ-域流一定是适应过程。不过由此产生的 $\mathcal{F}_t$ 不一定包含所有的零概率事件。为此，还需把所有的零概率事件添加到每个 $\mathcal{F}_t$ 中去。在离散模型中， σ-域流实际上就是证券的价格过程所形成的。但那时零概率事件与不可能事件是一回事，从而不引起上述问题。

如果允许随机过程的轨线是不连续的，那么我们也可定义可料过程以及其他一些过程。但是对于轨线是连续的随机过程来说，“前一个时刻”之类的说法就没有意义。因此，在我们以下讨论的模型中，策略也将是适应过程。当然，在更复杂的模型中，策略将仍是可料过程。

在前一讲的离散模型中，证券价格的变化在每一时刻都有两种可能。当时刻变化足够多，证券价格在固定时刻作为随机变量趋向于正态分布。而这种现象的连续化，就使证券价格的变化需要用所谓 Brown 运动来刻画。概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 上的随机过程 $(B_t)_{t \geqslant 0}$ 称为 **Brown 运动**是指：

1. (**连续性**)： (B_t) 的几乎所有的轨线是连续的；
2. (**增量的独立性**)：如果 $s < t$, 那么 $B_t - B_s$ 对 $\mathcal{F}_s = \sigma(B_u, u \leqslant s)$ 独立；

3. (**增量的平稳性**)：如果 $s < t$, 那么 $B_t - B_s$ 与 $B_{t-s} - B_0$ 有同样的概率分布。

可以证明，如果 $(B_t)_{t\geqslant 0}$ 是 Brown 运动，那么 $B_t - B_0$ 是均值为 μt, 方差为 $\sigma^2 t$ 的正态随机变量，其中 μ 和 σ 为常数。而对于 $0 \leqslant t_1 < \cdots < t_n$, $(B_{t_1}, \ldots, B_{t_n})$ 是正态随机向量。这里没有写出的证明的实质，正如我们在上一节中所指出的，犹如随机游走序列的项趋向于正态随机变量一样。

$B_0 = 0$ (在概率 1 的意义下 (a.s.)), $\mu = 0$, $\sigma = 1$ 的 Brown 运动称为标准 Brown 运动。以后我们一般总假定所讨论的 Brown 运动是标准 Brown 运动，同时，它还与某个 σ-域 (它甚至就是由该 Brown 运动所生成的) 联系在一起。这时，所谓 $\mathcal{F}_t$-Brown 运动就是一个适应过程，并且对于 $s < t$, $B_t - B_s$ 对 $\mathcal{F}_s$ 独立。

对于带 σ-域流 $(\mathcal{F}_t)_t \geqslant 0$ 的概率空间 $(\Omega, \mathcal{F}, \mathbf{P})$ 上的随机过程，我们同样可以定义条件数学期望，以及定义鞅。可积适应过程 $(M_t)_{t\geqslant 0}$ 称为**鞅** (这里"可积"意味着对于任何 t, $E[|M_t|] < +\infty$), 是指对于任何 $s \leqslant t$, $E[M_t|\mathcal{F}_s] = M_s$. 类似地也可定义上鞅和下鞅。对于连续鞅 (M_t) 来说，我们以后经常要利用下列重要的 **Doob 不等式**：

$$E\left[\sup_{0\leqslant t\leqslant T} |M_t|^2\right] \leqslant 4E[|M_T|^2].$$

其证明从略。

用标准 $\mathcal{F}_t$-Brown 运动可以生成好几种鞅；即我们有：

命题 10.1　设 $(B_t)_{t\geqslant 0}$ 是 $\mathcal{F}_t$-标准 Borwn 运动。那么

1. (B_t) 是 $\mathcal{F}_t$-鞅；
2. $(B_t^2 - t)$ 是 $\mathcal{F}_t$-鞅；
3. $\exp(\sigma B_t - (\sigma^2/2)t)$ 是 $\mathcal{F}_t$-鞅。

证明　1. 因为当 $s < t$ 时，$B_t - B_s$ 对 $\mathcal{F}_s$ 独立，故 $E[B_t - B_s|\mathcal{F}_s] = E[B_t - B_s] = 0$. 因此，$(B_t)$ 是 $\mathcal{F}_t$-鞅.

2. 对于 $s < t$, 由

$$\begin{aligned} E\left[B_t^2 - B_s^2|\mathcal{F}_s\right] &= E\left[(B_t - B_s)^2 + 2B_s(B_t - B_s)|\mathcal{F}_s\right) \\ &= E\left[(B_t - B_s)^2|\mathcal{F}_s\right] + 2B_s E[B_t - B_s|\mathcal{F}_s], \end{aligned}$$

以及由 (B_t) 是鞅，$E[B_t - B_s|\mathcal{F}_s] = 0$, 可得

$$E\left[B_t^2 - B_s^2|\mathcal{F}_s\right] = E\left[(B_t - B_s)^2|\mathcal{F}_s\right].$$

再由增量的平稳性和标准性，得到

$$E\left[(B_t - B_s)^2|\mathcal{F}_s\right] = E\left[B_{t-s}^2\right] = t - s.$$

因此， $E[B_t^2 - t|\mathcal{F}_s] = B_s^2 - s$.

3. 首先注意到，对于标准正态分布随机变量 g, 我们有

$$E\left[e^{\lambda}g\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{\lambda x}e^{-\frac{x^2}{2}}dx = e^{\frac{\lambda^2}{2}}.$$

同时，如果 $s < t$, 那么

$$E\left[e^{\sigma B_t} - \frac{\sigma^2 t}{2}\middle|\mathcal{F}_s\right] = e^{\sigma B_s - \frac{\sigma^2 t}{2}}E\left[e^{\sigma(B_t - B_s)}|\mathcal{F}_s\right],$$

并且由于 $B_t - B_s$ 对 $\mathcal{F}_s$ 独立，我们有

$$\begin{aligned} E\left[e^{\sigma(B_t-B_s)}|\mathcal{F}_s\right] &= E\left[e^{\sigma(B_t-B_s)}\right] &= E\left[e^{\sigma B_{t-s}}\right] \\ &= E\left[e^{\sigma g\sqrt{t-s}}\right] &= e^{\frac{\sigma^2(t-s)}{2}}. \end{aligned}$$

□

在离散情形下，我们知道证券组合的在自融资策略 ϕ 下的折现价值为

$$\tilde{V}_n(\phi) = V_0 + \sum_{j=1}^{n}\phi_j\Delta\tilde{S}_j.$$

把它连续化后，这一求和式将变为积分式 $\int_0^t \phi_s d\tilde{S}_s$. 这就涉及随机过程的积分。尤其是这里的折现价格将是由 Brown 运动生成的鞅，而人们早就知道 Brown 运动的几乎所有的轨线都是处处不可微的，以至这个积分的定义很难叙述清楚。从 1920 年代美国数学家 Wiener (由于 Wiener 的贡献，Brown 运动也称为 **Wiener 过程**) 的工作起，一直到 1940 年代日本数学家 Itô (伊藤清) 才算把随机积分的含义搞清。下面我们简述这方面的一些必要知识。其证明都从略。

设 $(B_t)_{t\geqslant 0}$ 是带 σ-域流的概率空间 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbf{P})$ 上的 $\mathcal{F}_t$-标准 Brown 运动。我们将对一类适应过程 (H_t) 定义下列形式的随机积分

$\int_0^t H_s dB_s$. 如同定义通常的 Lebesgue 积分那样，我们可先对“阶梯适应过程”(H_t) 来定义自然的积分。$(H_t)_{0\leqslant t\leqslant T}$ 称为**初等可料过程**，是指

$$H_t = \phi_i, \qquad t \in (t_{i-1}, t_i], \qquad 0 = t_0 < t_1 < \cdots < t_p = T,$$

其中 ϕ_i 是 $\mathcal{F}_{t_{i-1}}$-可测的有界随机变量。这里由于这样的初等过程不再是连续的，适应与可料就有所区别。对于这样的初等过程，当 $t \in (t_k, t_{k+1}]$ 时可以定义

$$\int_0^t H_s dB_s = \sum_{i=1}^{k} \phi_k (B_{t_i} - B_{t_{i-1}}) + \phi_{k+1}(B_t - B_{t_k}).$$

命题 10.2　如果 $(H_t)_{0\leqslant t\leqslant T}$ 为初等过程，那么

1. $\left(\int_0^t H_s dB_s\right)_{0\leqslant t\leqslant T}$ 是 $\mathcal{F}_t$-连续鞅；
2. $$E\left[\left(\int_0^t H_s dB_s\right)^2\right] = E\left[\int_0^t H_s^2 ds\right],$$
3. $$E\left[\sup_{t\leqslant T}\left|\int_0^t H_s dB_s\right|^2\right] \leqslant 4E\left[\int_0^T H_s^2 ds\right].$$

令

$$\mathcal{H} = \left\{(H_t)_{0\leqslant t\leqslant T} \quad \text{为对 } (\mathcal{F}_t)_{t\geqslant 0} \text{ 适应的过程，} \quad E\left[\int_0^T H_s^2 ds\right] < +\infty\right\}.$$

那么对于这一类适应过程，可以通过下列命题来定义随机积分：

命题 10.3　设 (B_t) 为 $\mathcal{F}_t$-标准 Brown 运动。那么存在唯一的 $\mathcal{H}$ 到定义在 $[0,T]$ 上的 $\mathcal{F}_t$-连续鞅的映射 J, 使得

1. 如果 $(H_t)_{t\leqslant T}$ 是初等过程，那么对于任何 $t \in [0,T]$,
$$J(H)_t = \int_0^t H_s dB_s \quad \text{a.s.};$$
2. 对于任何 $t \leqslant T$, $E\left[J(H)_t^2\right] = E\left[\int_0^t H_s^2 ds\right]$.

这里唯一性都是在概率 1 (a.s.) 的意义下确定的。

由此，我们就对 $\mathcal{H}$ 中的适应过程定义了随机积分 $\int_0^t H_s dB_s = J(H)_t$. 对于这样的随机积分，下列 Doob 不等式仍然成立：

$$E\left[\sup_{0\leqslant t\leqslant T}\left|\int_0^t H_s dB_s\right|^2\right] \leqslant 4E\left[\int_0^T H_s^2 ds\right].$$

对于另一类更广的适应过程：

$$\tilde{\mathcal{H}} = \left\{(H_t)_{0\leqslant t\leqslant T} \quad \text{为对 } (\mathcal{F}_t)_{t\geqslant 0} \text{ 适应的过程，} \quad \int_0^T H_s^2 ds < +\infty \ a.s.\right\}.$$

也能类似地定义随机积分，但其相应的 $(\int_0^t H_s dB_s)_{0\leqslant t\leqslant T}$ 不再一定是鞅，而只是连续随机过程。不过它有下列意义下的连续性：如果 $\{H^n\} \subset \tilde{\mathcal{H}}$ 满足 $\int_0^T (H_s^n)^2 ds$ 按概率趋于零，那么 $\sup_{0\leqslant t\leqslant T}|\int_0^t H_s^n dB_s|$ 也按概率趋于零。

有了这些准备以后，我们就可对下列形式的随机微分方程给出明确的数学定义：

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \qquad X_0 = Z.$$

如果 $\sigma = 0$, 这就是普通的微分方程，在力学上它刻画了每一时刻的运动速度 (大小和方向) 的变化。有了后一项后，这一运动速度受到了随机干扰。直观上似乎很好理解，但在数学上以前很难说得很确切。定义了随机积分以后，这样的随机微分方程可以通过所谓 Itô 过程来描述。

设 $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbf{P})$ 为一带 σ-域流的概率空间。$(B_t)_{t\geqslant 0}$ 为 $\mathcal{F}_t$-Brown 运动。那么下列形式的实值随机过程 $(X_t)_{0\leqslant t\leqslant T}$ 称为 **Itô 过程**：

$$\forall t \in [0, T], \qquad X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dB_s \quad \text{a.s.},$$

其中 X_0 为 $\mathcal{F}_0$- 可测；(K_t) 和 (H_t) 为对 $\mathcal{F}_t$ 的适应过程，且满足

$$\int_0^T |K_s| ds < +\infty \quad \text{a.s.}, \qquad \int_0^T H_s^2 ds < +\infty \quad \text{a.s.}.$$

可以证明，下列命题成立：

命题 10.4　设 $(M_t)_{0\leqslant t\leqslant T}$ 是连续鞅，且

$$M_t = \int_0^t K_s ds, \quad \int_0^T |K_s| ds < +\infty \quad \text{a.s.},$$

那么

$$\forall t \leqslant T, \qquad M_t = 0 \quad \text{a.s.}.$$

正是由于有这一命题，Itô 过程在几乎处处的意义下有唯一表示；即，如果

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dB_s = X_0' + \int_0^t K_s' ds + \int_0^t H_s' dB_s,$$

那么由于 X_0, X_0' 和两个随机积分项都是连续鞅，可先导出几乎处处有 $K_s = K_s'$, 再导出 $X_0 = X_0' \quad \text{a.s.}$. 最后，再得到几乎处处有 $H_s = H_s'$.

Itô 过程使前面提到的随机微分方程有了确切的含义。Itô 证明，如果

$$\int_0^T |b(s, X_s)| ds < +\infty \quad \text{a.s.}, \qquad \int_0^T |\sigma(s, X_s)|^2 ds < +\infty \quad \text{a.s.},$$

$x \mapsto b(t,x)$ 和 $x \mapsto \sigma(t,x)$ 是 Lipschitz 函数， b 和 σ 有界，以及 $E[Z^2] < +\infty$, 那么在几乎处处的意义下，该随机微分方程有唯一解 $(X_t)_{0\leqslant t\leqslant T}$, 且 $E\left[\sup_{0\leqslant s\leqslant T} |X_s|^2\right] < +\infty$.

随机微分方程的研究还引起了随机分析中极为重要的 Itô 公式。这个公式告诉我们，对于 Itô 过程来说，普通的复合函数求导公式不再成立。举一个最简单的例子来看，设 (B_t) 是标准 Brown 运动。如果对它应用普通的复合函数求导公式：$f(t)^2 = f(0)^2 + 2\int_0^t f(s)df(s)$, 那么应该有 $B_t^2 = 2\int_0^t B_s dB_s$. 但这是不可能的，因为右端是鞅，其均值为零；而左端的均值是 t. 正确的求导公式由下列 **Itô 公式**给出：

定理 10.1　设 $(X_t)_{0\leqslant t\leqslant T}$ 是 Itô 过程:

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dB_s,$$

$f(\cdot)$ 是二次连续可微函数，那么

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2}\int_0^t f''(X_s) d\langle X, X\rangle_s,$$

其中

$$\langle X, X\rangle_t = \int_0^t H_s^2 ds,$$

$$\int_0^t f'(X_s) dX_s = \int_0^t f'(X_s) K_s ds + \int_0^t f'(X_s) H_s dB_s.$$

如果 $t \mapsto f(t,x)$ 一次连续可微， $x \mapsto f(t,x)$ 二次连续可微，那么更一般地有

$$f(t,X_t) = f(0,X_0) + \int_0^t f'_s(s,X_s)ds + \int_0^t f'_x(s,X_s)dX_s + \frac{1}{2}\int_0^t f''_{xx}(s,X_s)d\langle X,X\rangle_s.$$

我们不在这里给出这一公式的证明，但是从形式上来看，我们在上一节已经指出如何来理解这一公式。

以上的讨论都是一维的。对于向量随机过程，容易推广上面的所有讨论。

10.3　现在我们来讨论经典的 Black–Scholes 模型。在这个模型中，时间变量是连续的，且只有两种证券。第一种证券是无风险证券，相当于银行帐户。其依赖于时间 t 的价格 S_t^0 满足下列 (常) 微分方程：

$$dS_t^0 = rS_t^0 dt,$$

其中 r 是“瞬时”银行利率，假设它是正常数。不妨取 $S_0^0 = 1$, 那么 $S_t^0 = e^{rt}$. 另一种证券是有风险证券。它的价格满足下列随机微分方程：

$$dS_t = S_t(\mu dt + \sigma dB_t),$$

其中 μ 和 σ 是常数， (B_t) 是标准 Brown 运动。我们将把问题限制在有限时间区间 $[0,T]$ 上来研究，其中 T 是 (欧式) 期权到期的时刻。正如上面的讨论中所述，这一方程的解可表示为

$$S_t = S_0 \exp\left(\mu t - \frac{\sigma^2}{2}t + \sigma B_t\right),$$

其中 S_0 为该证券的初始价格。这样，对于固定的 t, S_t 是对数正态分布随机变量，即， $\log S_t$ 是正态分布随机变量。更确切地说， $(\log S_t)$ 是 Brown 运动 (不一定是标准的)。

对这样的市场模型，我们同样可以定义**策略** $\phi = (\phi_t)_{0\leqslant t\leqslant T} = ((H_t^0, H_t))$, 它是一个二维的 (对 Brown 运动 (B_t) 生成的 σ-域流 $\mathcal{F}_t$) 的适应过程，其中 H_t^0 和 H_t 分别表示在时刻 t 的对两种证券的交易量。它所形成的证券组合在时刻 t 的价值为

$$V_t(\phi) = H_t^0 S_t^0 + H_t S_t.$$

策略 ϕ 为**自融资**的条件现在变为：

$$dV_t(\phi) = H_t^0 dS_t^0 + H_t dS_t.$$

但是为了使这个等式有意义，以至 $(V_t(\phi))$ 成为 Itô 过程，需假设

$$\int_0^T |H_t^0| dt < +\infty \quad \text{a.s.}, \qquad \int_0^T H_t^2 dt < +\infty \quad \text{a.s.}.$$

这样，积分

$$\int_0^T H_t^0 dS_t^0 = \int_0^T H_t^0 r e^{rt} dt$$

和随机积分

$$\int_0^T H_t dS_t = \int_0^T (H_t S_t \mu) dt + \int_0^T \sigma H_t S_t dB_t$$

都有定义。因此，策略 $\phi = ((H_t^0, H_t))$ 为自融资的确切定义应该是下列两个条件：

1. $\int_0^T |H_t^0| dt + \int_0^T H_t^2 dt < +\infty \quad \text{a.s.}$,
2. 对于任何 $t \in [0, T]$, 有

$$H_t^0 S_t^0 + H_t S_t = H_0^0 S_0^0 + H_0 S_0 + \int_0^t H_u^0 dS_u^0 + \int_0^t H_u dS_u \quad \text{a.s.}.$$

与以前一样，我们以 $\tilde{S}_t = e^{-rt} S_t$ 表示风险证券的折现价格。那么我们同样有下列命题：

命题 10.5　设 $\phi = ((H_t^0, H_t))_{0 \leqslant t \leqslant T}$ 为二维适应过程，且满足 $\int_0^T |H_t^0| dt + \int_0^T H_t^2 dt < +\infty \quad \text{a.s.}$. $V_t(\phi) = H_t^0 S_t^0 + H_t S_t$, $\tilde{V}_t(\phi) = e^{-rt} V_t(\phi)$. 那么 ϕ 为自融资策略当且仅当

$$\forall t \in [0, T], \qquad \tilde{V}_t(\phi) = V_0(\phi) + \int_0^t H_u d\tilde{S}_u \quad \text{a.s.}.$$

我们已经知道，$(\tilde{S}_t) = (e^{-rt} S_t)$ 一般不一定是鞅。如果遵循离散模型中的思路，我们的目标将是求出某种等价概率测度，使得 $(\tilde{S}_t)$ 是鞅。于是由此就可对期权以至一般的衍生证券来定价。对于一般的可测空间，两种概率测度等价是指它们互相对对方绝对连续，或者可以说，对这两种概率测度来说，它们的零概率事件是一样的。而要把 $(\tilde{S}_t)$ 这样的形式的适应过程变换成一个 Brown 运动，需要应用下列重要的 Girsanov 定理：

定理 10.2 设 $(\theta_t)_{0\leqslant t\leqslant T}$ 为满足 $\int_0^T \theta_s^2 ds < \infty$ a.s. 的适应过程，且如下定义的 $(L_t)_{0\leqslant t\leqslant T}$ 是鞅:

$$L_t = \exp\left(-\int_0^t \theta_s dB_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right).$$

那么相对于 L_T 的概率密度所定义的概率测度 $\mathbf{P}^{(L)}$ 来说，定义为 $W_t = B_t + \int_0^t \theta_s ds$ 的过程 $(W_t)_{0\leqslant t\leqslant T}$ 是标准 Brown 运动。

由于

$$d\tilde{S}_t = -re^{-rt}S_t dt + e^{-rt}dS_t = \tilde{S}_t\left((\mu - r)dt + \sigma dB_t\right),$$

我们可在这个定理中取 $\theta_t = (\mu - r)/\sigma$, 那么相对于这个 θ 的 (L_t) 显然是鞅。从而，令 $W_t = B_t + (\mu - r)t/\sigma$, 那么它就对某个等价概率测度 $\mathbf{P}^*$ 为标准 Brown 运动，并且

$$d\tilde{S}_t = \tilde{S}_t \sigma dW_t,$$

$$\tilde{S}_t = \tilde{S}_0 \exp(\sigma W_t - \sigma^2 t/2),$$

即 $(\tilde{S}_t)$ 对 $\mathbf{P}^*$ 也是鞅。

有了这个等价概率测度以后，我们就可对 (欧式) 期权 (未定权益) 定价问题给出回答。这里期权定义为非负的 $\mathcal{F}_T$-可测的随机变量，并且还要求对上面定义的概率 $\mathbf{P}^*$ 平方可积。原因在于在这样的条件下，可以证明，每个这样的期权 h 都可以用某个自融资策略 ϕ 来达到。于是 h 在时刻 t 的定价就可通过

$$V_t = E^*\left[e^{-r(T-t)}h|\mathcal{F}_t\right]$$

来给出。尤其是，如果 $h = f(S_T)$, 那么

$$\begin{aligned} V_t &= E^*\left[e^{-r(T-t)}f(S_T)|\mathcal{F}_t\right] \\ &= E^*\left[e^{-r(T-t)}f\left(S_t e^{r(T-t)}e^{\sigma(W_T - W_t) - (\sigma^2/2)(T-t)}\right]\middle|\mathcal{F}_t\right). \end{aligned}$$

因为 S_t 是 $\mathcal{F}$-可测的，且在概率测度 $\mathbf{P}^*$ 下， $W_T - W_t$ 对 $\mathcal{F}_t$ 独立，记 $V_t = F(t, S_t)$, 我们得到

$$\begin{aligned} F(t,x) &= E^*\left[e^{-r(T-t)}f\left(xe^{r(T-t)}e^{\sigma(W_T - W_t) - (\sigma^2/2)(T-t)}\right)\right] \\ &= \frac{e^{-r(T-t)}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f\left(xe^{(r-(\sigma^2/2))(T-t)+\sigma y\sqrt{T-t}}\right)e^{-\frac{y^2}{2}}dy. \end{aligned}$$

对于欧式买入期权，$f(x)=(x-K)_+$, 由此式就能计算得到其价格 $c(t,x)$ 为：

$$c(t,x)=xN(d_1)-Ke^{-r(T-t)}N(d_2),$$

其中

$$d_1=\frac{\log(\frac{x}{K})+(r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}},\qquad d_2=d_1-\sigma\sqrt{T-t}.$$

$$N(d)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d}e^{-y^2/2}dy.$$

而对于欧式卖出期权， $f(x)=(K-x)_+$, 由此可得到它的价格 $p(t,x)$ 为：

$$p(t,x)=Ke^{-r(T-t)}N(-d_2)-xN(-d_1).$$

它们就是经典的 Black–Scholes 公式。

以上我们通过与离散模型的类似想法，导得了经典的 Black–Scholes 公式。如果读者对随机分析比较熟悉，这里的数学推导似乎没有几步。但是对比前一讲中的讨论，我们可以发现，如果要建立系统的连续模型的定价理论，我们还需要建立无套利、可生存市场、完全市场等一系列概念，以及有关的一系列定理，才能使这里的经济思想得到比较完整的叙述。

此外， Black–Scholes 公式还可应用于许多其他不同于欧式股票期权的情形。其中有一些情形都不需要修改公式。例如，对于长期零息债券的短期期权，可以原封不动地照搬原来的公式，只是把其中的股票价格改为债券价格。但是有些情形需要对原公式稍作修改。例如，如果股票是有分红的，并且假定其分红以连续分红率 d 来执行，那么相应的 Black–Scholes 公式应该把分红的价值考虑在内。这时公式变为：

$$c(t,x)=xe^{-d(T-t)}N(d_1)-Ke^{-r(T-t)}N(d_2),$$

其中

$$d_1=\frac{\log(\frac{xe^{-d(T-t)}}{K})+(r+\frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}},\qquad d_2=d_1-\sigma\sqrt{T-t}.$$

$$p(t,x)=Ke^{-r(T-t)}N(-d_2)-xe^{-d(T-t)}N(-d_1).$$

这一公式也经常用于指数期权情形。这时，可把指数看作一个有连续分红的股票。还可以用于外汇期权的情形。这时，d 是外汇利率，x 是汇率。如此等等。

10.4　然而，在经典的 Black–Scholes 的原始论文中，他们走的又是另一条路。其中更多的是利用了偏微分方程作为工具。

他们的办法是先为 (欧式买入) 期权价格导出一个随机微分方程。事实上，由 Itô 公式，立即可得

$$dc(t,S_t)=\left(c_t(t,S_t)+c_x(t,S_t)\mu S_t+\frac{1}{2}c_{xx}(t,S_t)\sigma^2 S_t\right)dt$$

$$+c_x(t,S_t)\sigma S_t dB(t),\qquad c(T,S_T)=(S_T-K)_+.$$

注意到 dS_t 的定义，它也可写成

$$dc(t,S_t)=c_x(t,S_t)dS_t+c_t(t,S_t)dt+\frac{1}{2}c_{xx}(t,S_t)\sigma^2 S_t^2 dt,$$

$$c(T,S_T)=(S_T-K)_+.$$

这是个从终端时间 T 出发来求出 T 前面的时刻 t 的所谓“倒向随机微分方程”。不过这样的倒向方程无法求解，因为我们还没有把它与无风险证券相联系，不能计算它的折现价值。其次， Black 和 Scholes 从考虑期权有减少风险的作用出发而认为，在时刻 t 以价格为 S_t 卖出一份股票，必须同时买进 $1/c_x(t,S_t)$ 份期权来保值。这是因为当股价变化 dS_t 时，期权价格就近似地变化 $c_x(t,S_t)dS_t$. 两项变化正好相抵。卖出 $c_x(t,x(t))$ 份股票，买进一份期权的情况也一样。于是时价为

$$\beta(t)=c_x(t,S_t)S_t-c(t,S_t)$$

其价格变化为

$$d\beta_t=c_x(t,S_t)dS_t-dc(t,S_t)$$

的证券是一项无风险证券，从而它应满足常微分方程

$$d\beta_t=\beta_t r dt.$$

由此再把 S_t 代替为 x, 就可导得，

$$c_t(t,x)+\frac{1}{2}c_{xx}(t,x)\sigma^2x^2=-rc_x(t,x)x+rc(t,x).$$

这是一个以 t 和 x 为变量的偏微分方程，并且可以经过变换后化为熟知的扩散方程。再加上终端条件 $c(T,x)=(x-K)_+$, 它就可以求解，并且同样得到上面的 Black–Scholes 公式。

如果我们用 $u=u(t,x)$ 来代替 $c(t,x)$ 作为一般的欧式股票衍生证券的定价函数，并且把偏导数记号写得更明确，那么我们可以发现，下

列所谓 **Black–Scholes 方程**对于任何欧式股票衍生证券的定价都是成立的：

$$\frac{\partial u(t,x)}{\partial t}+\frac{\sigma^2x^2}{2}\frac{\partial^2 u(t,x)}{\partial x^2}+rx\frac{\partial u(t,x)}{\partial x}-ru(t,x)=0.$$

所不同的仅仅是终端条件。由此，我们就可以用偏微分方程方法求得各种欧式股票期权的定价 (参看附录)。

除了欧式期权以外，还有许多其他形式的期权。其中最常见的是美式期权。所谓美式期权是到期日以前都可以执行的期权。美式期权的定价比欧式期权的定价要复杂得多。但是原理是一样的。从偏微分方程的角度来说，它所应该满足的方程仍然是 Black–Scholes 方程，但是终端条件不再一样。由于它并不一定到期执行，从而其执行时间也是需要确定的。这种类型的偏微分方程问题称为自由边界问题，其求解方法比较复杂。然而，前一讲中提出的二叉树方法仍然在这种情况下可继续应用。目前，二叉树方法与偏微分方程的计算方法之间的关系，也是与此有关的一个受到重视的研究课题。

Black–Scholes 期权定价模型只是连续时间金融学中的一个典型论题。事实上，我们前面讨论过的所有论题(资产定价基本定理、Modigliani–Miller 定理、证券组合选择、CAPM、一般经济均衡等等) 都可以有连续时间的框架。有兴趣的读者可参阅 Merton 的原著 [15]。下面，我们将以同样的思路来用连续时间模型来讨论利率的期限结构。

10.5 直到现在为止，我们讨论中所牵涉到的利率作为无风险 (净) 收益率，它在模型中是个不变的常数。利率是作为某种货币计量基准来引入的。由于期权定价等讨论所涉及的时间段相对较短 (一般至多几个月)，故把利率作为常数处理问题不大。但是现实世界中的银行利率、债券利率是随时间变化而变化的。因此，在考虑时间较长的金融产品定价问题就不能再以一项不变的无风险收益率来作基准。所谓**利率的期限结构** (term structure of interest rates) 就是指利率与期限的关系。就如我们通常所理解的那样，一般长期利率要比短期利率要高 (在某些反常情况下也会相反)。这种利率与期限的关系通过**收益曲线** (yield curve) $y=y(s)$ 来描述，例如，在美国，每天都要公布国库券的收益曲线，即，3 个月、6 个月、1 年、2 年直至 30 年的国库券利率值。它就经常作为各种金融产品定价的基准。前面讨论的期权定价问题在原则上都可推广到利率也有随机变化的情形。但是这节中，我们要讨论的是：当利率有随机变化时，怎样来为债券定价。

为了理解怎样用收益曲线来为债券定价，我们先来讨论一个简单的债券定价问题。世上债券种类五花八门。但是基本上可分为到期付本还息和分期付本还息两种。这样，理论上，广义的**付息债券** (coupon bond) 可以看作最基本的债券，它的特点是分期付本还息或者分期付息。到期付本还息的债券可以看作它的特例：**零息债券** (zero-coupon bond)。基准收益曲线 $y = y(s)$ 中的自变量 s 可看作零息债券的期限，y 是相应的利率，其中利率的单位都被折算为用统一的单位期限增益来计算。于是，如果 $s = 0, 1, \ldots, T$, 那么 $s = 0$ 时的一元钱，到 $s = 1$ 时，它将变为 $(1+y(1))$ 元，到 $s = 2$ 时，它将变为 $(1+y(2))^2$ 元，……，一般情况下，到 s 时，它将变为 $(1+y(s))^s$ 元。同样，$s = 1$ 时的一元钱，将等于 $s = 0$ 时的 $1/(1+y(1))$ 元， $s = 2$ 时的一元钱，将等于 $s = 0$ 时的 $1/(1+y(2))^2$ 元，……，一般情况下， s 时的一元钱，将等于 $s = 0$ 时的 $1/(1+y(s))^s$ 元。这相当于**折现债券** (discount bond, 到期付票面价值的债券) 的发行价计算。假定 $y(s)$ 都不是随机变量，那么我们也可把它看作“未来状态”为时刻 $s = 1, 2, \ldots, T$ 的随机变量。这时，同样在无套利假设下， $1/(1+y(1)), 1/(1+y(2))^2, \ldots 1/(1+y(s))^s, \ldots, 1/(1+y(T))^T$ 就都可以看作各个“状态价格”。由此就可对任何债券定价。例如，某种付息债券 $\bar{x}$ 在 $s = T$ 时到期，那时债券持有者将得到 K 元，但在 $s = 1, 2 \ldots, T-1$ 元，债券持有者还能每次得到 (利息) R 元。那么 x 的当前价格为

$$p(\bar{x}) = \sum_{s=1}^{T-1} \frac{R}{(1+y(s))^s} + \frac{K}{(1+y(T))^T}.$$

更一般的在 T 到期的付息债券可用 T 维向量 $x = (x_1, x_2, \ldots, x_T)$ 来表示。那么在基准收益曲线 $y = y(s)$ 下，其当前价格为

$$p(x) = \sum_{s=1}^{T} \frac{x_s}{(1+y(s))^s}.$$

我们可以发现常见的论著中的所有有关债券、股票通过利率来简单定价的公式都是这个公式的特例。这里我们甚至还可考虑税收等因素。不难看出，这一公式还可推广到 $y(s)$ 为有限状态的随机变量的情形。

利率期限结构的理论研究目前一般用连续时间模型来讨论。假定 (连续复合即时) 利率 $r(t)$ 随时间 t 变化。那么债券的价值随时间的变化 $V = V(t)$ 应该满足下列微分方程：

$$\frac{dV(t)}{dt} = r(t)V(t), \qquad V(0) = V_0.$$

由此可得

$$V(t) = V_0 e^{\int_0^t r(s)ds}.$$

如果考虑的是折现债券，已知的是到期值 $V(T) = V_T$, 那么将有

$$V(t) = V_T e^{-\int_t^T r(s)ds}.$$

上述收益曲线通常是对平均利率而言的。记由 $t(<T)$ 到 T 的平均利率为 $R(t,T)$, 即

$$V_T = V(t)e^{(T-t)R(t,T)}.$$

那么比较上述两式可得

$$R(t,T) = \frac{1}{T-t}\int_t^T r(s)ds.$$

令 $s = T-t$. 那么 $y = y(s) = R(t,T)$ 就是前面所说的收益曲线。有时我们就取 $V_T = 1$. 这时记 $V(t) = P(t,T) = e^{-\int_t^T r(s)ds}$ 就可看作折现债券在时刻 t 的价格。这时将有

$$y(T-t) = R(t,T) = -\frac{\log P(t,T)}{T-t}.$$

然而，利率 $r(t)$ 一般随时间随机变化。通常假定它是一个 $\mathcal{F}_t$-适应过程，并满足下列形式的随机微分方程：

$$dr(t) = u(t,r(t))dt + w(t,r(t))dB_t, \tag{7}$$

其中 B_t 为标准 Brown 运动，u 和 w 为 (t,r) 的确定函数。在这种情况下，债券的价值不但依赖于时间，还依赖于利率的随机变化，即 $V = V(t,r(t))$. 为给出 V 的表达式，一种办法是利用无套利假设给出一个对于债券市场的等价概率鞅测度 E^*, 并取

$$R(t,T) = E^*\left[\frac{1}{T-t}\int_t^T r(s)ds\,\middle|\,\mathcal{F}_t\right].$$

由此就可导得 $V(t,r(t)) = V_T e^{(T-t)R(t,T)}$.

另一种办法是利用风险对冲。但是与期权定价问题根本不同之处在于 r 并非是 V 的标的资产，V 也并非用来对冲 r 的风险。为对冲风险，我们应该用两种不同期限的债券的适当组合来对冲风险。假设有两种到期日分别是 T_1 和 T_2 的债券。前者的价格 V_1, 后者的价格为 V_2. 我们用如下的组合来套期保值：

$$\Pi = V_1 - \Delta V_2,$$

其中 Δ 是一个适当的数。这一组合在时间 dT 中的变化为

$$d\Pi = \frac{\partial V_1}{\partial t}dt + \frac{\partial V_1}{\partial r}dr + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2}dt - \Delta\left(\frac{\partial V_2}{\partial t}dt + \frac{\partial V_2}{\partial r}dr + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}dt\right),$$

其中对 r 和 t 的函数应用了 Itô 公式。由此可以看到取

$$\Delta = \frac{\partial V_1/\partial r}{\partial V_2/\partial r},$$

就能消除 $d\Pi$ 中的随机因素，使得

$$\begin{aligned} d\Pi &= \left(\frac{\partial V_1}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2} - \frac{\partial V_1/\partial r}{\partial V_2/\partial r}\left(\frac{\partial V_2}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}\right)\right) \\ &= r\Pi dt = r\left(V_1 - \frac{\partial V_1/\partial r}{\partial V_2/\partial r}V_2\right)dt, \end{aligned}$$

其中我们利用了无风险利率为 r 的事实。这样，我们得到偏微分方程

$$\left(\frac{\partial V_1}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2} - \frac{\partial V_1/\partial r}{\partial V_2/\partial r}\left(\frac{\partial V_2}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}\right)\right) = r\left(V_1 - \frac{\partial V_1/\partial r}{\partial V_2/\partial r}V_2\right).$$

把带 V_1 的项放在左端，把带 V_2 的项放在右端，由此可得

$$\left(\frac{\partial V_1}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2} - rV_1\right)\Big/\frac{\partial V_1}{\partial r} = \left(\frac{\partial V_2}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2} - rV_2\right)\Big/\frac{\partial V_2}{\partial r}.$$

这是一个方程，两个未知函数 V_1 和 V_2. 但是它们分别在两端，因此两端都应等于与 V 无关的函数 $a(t,r)$. 为今后应用的方便，不妨记

$$a(t,r) = w(t,r)\lambda(t,r) - u(t,r),$$

由此即得**零息债券定价方程：**

$$\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + (u - \lambda w)\frac{\partial V}{\partial r} - rV = 0. \qquad (8)$$

其终端条件为 $V(r,T) = V_T$, 但边界条件要取决于 u 和 w 的形式。

这里的 λ 有其特殊意义。事实上，由 Itô 公式，

$$dV = w\frac{\partial V}{\partial r}dB_t + \left(\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + u\frac{\partial V}{\partial r}\right)dt.$$

从而由 (8) 可得

$$dV = w\frac{\partial V}{\partial r}dB_t + \left(w\lambda\frac{\partial V}{\partial r} + rV\right)dt,$$

或

$$dV - rVdt = w\frac{\partial V}{\partial r}(dB_t + \lambda dt).$$

这一等式说明，零息债券并非是无风险证券，因为无风险证券的右端应该是零。右端可以解释为承担风险的超额收益，而函数 λ 通常称为**风险的市场价格**，它是类似于 Sharpe 比那样的量。这可以通过把方程 (8) 与 Black–Scholes 方程相比较来看出。在 Black–Scholes 方程中与 $u-\lambda w$ 相当的是无风险利率 r 与股价 x 的乘积，u 相当于 μx, w 相当于 σx. 于是 λ 就相当于 $(\mu-r)/\sigma$. 这恰好就是第五讲中提到的 Sharpe 比。

以上是对一般的利率随机微分方程 (7) 来讨论的。与期权定价理论不同的是，利率随机微分方程一般都不是“常收益、常波动率”那样的形式。文献中研究较深入的一些利率随机微分方程可在 (7) 的记号下，统一记成下列形式：

$$w(t,r)=\sqrt{\alpha(t)r-\beta(t)} \tag{9}$$

$$u(t,r)=\left(-\gamma(t)r+\delta(t)+\lambda(t,r)\sqrt{\alpha(t)r-\beta(t)}\right). \tag{10}$$

当 $\alpha=0$ 且所有参数都是常数时，可得

$$dr(t)=a(b-r(t))dt+\sigma dB_t,$$

其中 a,b,σ 都是常数。这是 Vasicek [15] 于 1977 年提出的模型。这一方程在没有 Brwon 运动干扰时，其解为 $r(t)=r(0)e^{-at}+b(1-e^{-at})$, 即随着时间趋向于无限大，利率趋向于常数 b. 而 a 可理解为利率的调节速度。在有 Brown 运动干扰时，满足这样的随机微分方程的随机过程是所谓 Ornstein–Uhlenback 过程的一个变形。它可以写成随机积分的形式：

$$r(t)=r(0)e^{-at}+b(1-e^{-at})+\sigma e^{-at}\int_0^t e^{as}dB_t.$$

由此可得方差 $\mathrm{Var}[r(t)]=\sigma^2(1-e^{-2at})/(2a)$.

当 $\beta=0$ 且所有参数都是常数时，可得

$$dr(t)=a(b-r(t))dt+\sigma\sqrt{r(t)}dB_t.$$

它与 Vasicek 模型的区别仅在于多了 $\sqrt{r(t)}$ 项。这是 Cox, Ingersoll 和 Ross [3] 于 1985 年提出的。这个模型几乎是目前最有名的利率期限结构的模型。常称 CIR 模型。加入 $\sqrt{r(t)}$ 的目的在于消除 Vasicek 模型中 $r(t)$ 有可能取负值的不合理情形。

对于这两种模型，有可能把 $R(t,T)$ 的表达式写出来。事实上，设

$$R(t,T)=\frac{1}{T-t}[B(t,T)r(t)-A(t,T)].$$

那么可以求得，对于 Vasicek 模型，

$$B(t,T)=\frac{1-e^{-a(T-t)}}{a},$$

$$A(t,T)=\frac{(B(t,T)-tT)(a^2b-\sigma^2/2)}{a^2}-\frac{\sigma^2B(t,T)^2}{4a}.$$

而对于 Cox–Ingersoll–Ross 模型，

$$B(t,T)=\frac{2(e^{\gamma(T-t)}-1)}{(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma},$$

$$A(t,T)=\frac{2ab}{\sigma^2}\log\left[\frac{2\gamma e^{(a+\gamma)(T-t)/2}}{(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma}\right],$$

其中 $\gamma=\sqrt{a^2+2\sigma^2}$.

上述两种模型的系数都依赖于 t 的情形是 Hull 与 White [9] 在 1990 年建议的。这时将得

$$dr(t)=a(t)(b(t)-r(t))dt+\sigma(t)dB_t,$$

$$dr(t)=a(t)(b(t)-r(t))dt+\sigma(t)\sqrt{r(t)}dB_t.$$

对于它们也能通过类似的方法得到相应的表达式。至于一般的讨论可参看 Duffie [5]。

另一类利率期限结构模型是从**远期利率**出发的。在离散情形下，时刻 t 的远期利率 $f(t)$ 定义为

$$1+f(t)=\frac{(1+y(t+1))^{t+1}}{(1+y(t))^t}.$$

这一定义的近似连续表达式为

$$e^{\Delta Tf(t,T)}=\frac{e^{\int_t^{T+\Delta T}r(s)ds}}{e^{\int_t^T r(s)ds}},$$

从而在连续情形下，时刻 t 的远期利率 $f(t,T)$ 定义为

$$f(t,T)=\lim_{\Delta T\to0}\frac{\displaystyle\int_t^{T+\Delta T}r(s)ds-\int_t^T r(s)ds}{\Delta T}=-\frac{\partial\log P(t,T)}{\partial T}.$$

它与即时利率 $r(t)$ 的关系为 $r(t)=f(t,t)$. 但是这都是在非随机的情形下。在随机情形下，上面的一些模型都是对 r 提出随机模型。1987 年，

Heath, Jarrow 和 Morton [6] 建议对远期利率 $f(t,T)$ 提出随机模型，即假定远期利率满足下列形式的随机微分方程：

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dB_t,$$

其中 Brown 运动项 dB_t 还可假定为多维的。这一模型的好处在于，随着 T 的不同可得到不同的结果。由此得出的收益曲线比单纯从同一个即时利率 r 的随机模型得出的收益曲线要合理得多。当然，其代价是数学上也要复杂得多。

此外，以利率作为标的资产的衍生证券种类也很多。例如，债券期权，掉期 (swap, 两种不同期限的债券的互换)等等。对它们的定价仍然可以用与以前一样的方法。类似的讨论还经常对汇率进行。

说明 虽然在 Black–Scholes 理论以前，一直可追溯到 1900 年的 Bachelier 的学位论文，人们已经把金融经济学在时间上连续化，但是正式称得上“连续时间金融学”的标志，还应该从 Black–Scholes 的经典论文 [2] 和 Merton 的经典论文 [15] 于 1973 年发表算起。当时，Black 与 Scholes 早写好他们的论文，但迟迟不能发表 (据说，后来的发表曾通过 Miller 去说服编辑部)。在此期中他们又与 Merton 进一步讨论，使 Merton 能为此建立更一般的框架。最后，两篇论文几乎同时发表。

“连续时间金融学”必须以随机分析作为数学工具。这成了人们学习这一学科领域的拦路虎。Huang–Litzenberger [7] 和 Jarrow [11] 对此基本上采取回避态度。Ingersoll [10] 和 Hull [8] 则用描述性的方式来介绍必要的数学工具。数学上比较完整的叙述是 Duffie [4], [5]，但对数学基础不足的读者来说，这两本专著都不太好读。比较好读的书，正如我们以前已经介绍过的，是 Baxter–Rennie [1]。我们的叙述则仍比较接近于 Lamberton–Lapeyre [12]。我们希望我们的叙述既比较容易接受，又不失数学上的严谨。

其实连续时间金融学所需要的随机分析，最重要的还是一些基本概念：Brown 运动、鞅、随机积分、随机微分方程、Itô 公式等。对这些概念的一般理解并不那么可怕。只有对这些概念有一定的理解，才能明白连续时间金融学究竟在说些什么。至于许多技术细节，对一般的读者来说并不必要。

从经济思想来说，这里最根本的还是资产定价基本定理。深入讨论时主要是两种方法：求出等价概率鞅测度和求出 (连续的) 无风险证券组合。前者理论上涉及一条所谓 Girsanov 定理，实际应用时会涉及许多概率运算，可采取 Monte Carlo 模拟计算方法。后者经常把问题归结为解偏微分方程，可通过偏微分方程的数值方法来求解。但是前一讲中的二叉树方法对于连续时间金融学来说，无论是理论上还是实际计算上，始终都是有意义的。

利率期限结构问题在连续时间金融学中几乎与衍生证券定价问题一样重要。但在 Huang–Litzenberger [7] 和 Jarrow [11] 中几乎没有提。Ingersoll [10] 中也讲得不多。倒是面向实用的 Hull [8] 讲得不少。利率期限结构问题主要是要为利率的随机变化建立模型。这对债券定价十分重要。

思考题和习题

1. 什么是随机游走？什么是 Brown 运动？什么是白噪声？
2. 在 Black–Scholes 模型中是怎样对股票价格作出假定的？它的经济意义是什么？
3. 什么是随机微分方程？什么是 Itô 公式？怎样理解随机微分方程与 Itô 公式？
4. Black–Scholes 模型的出发点是什么？怎样对折现价格求出等价概率鞅测度？它是利用什么数学定理来求得的？
5. 怎样利用股票与期权的无风险组合来导出 Black–Scholes 方程？怎样在其中利用 Itô 公式？
6. 什么是零息债券定价方程？它与 Black–Scholes 方程有什么相同和不同之处？
7. 什么是利率随机微分方程模型？常用的模型有哪些？

参考文献

[1] Baxter, M., and Rennie, 1996, A., Financial Calculus, An Introduction to Derivative Pricing, Cambridge University Press, Cambridge.

[2] Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy*, **81:** 637–654.

[3] Cox, J. C., Ingersoll, J. E. Jr., and Ross, S. A., 1985, A theory of the term structure of interest rates, *Econometrica*, 53: 385–408.

[4] Duffie, D., 1988, Security Markets: Stochastic Models, Academic Press, Boston.

[5] Duffie, D., 1997, Dynamic Asset Pricing Theory, Second ed., Princeton University Press, Princeton, N. J..

[6] Heath, D., R. Jarrow and A. Morton, 1992, Bond pricing and the term structure of interest rates, *Economerica*, **60**: 77–106.

[7] Huang, Chi-fu & R. H. Litzenberger, 1988, Foundations for Financial Economics, Prentice–Hall, Englewood Cliffs, N. J..

[8] Hull, J. C., 1999, Options, Futures, and Other Derivatives, 4th ed., Prentice–Hall, N.J.. (第三版有中译本：赫尔，J., 2000，期权、期货和衍生证券，张陶伟译，华夏出版社).

[9] Hull, J. and White, A., 1990, Pricing interest-rate-derivatives securities, *Review of Financial Studies*, 3: 573–592.

[10] Ingersoll, J. E. Jr., 1987, Theory of Financial Decision Making, Rowman & Littlefield Publishers, Maryland.

[11] Jarrow, R. A., 1988, Finance Theory, Prentice-Hall, Englewood Cliffs, N. J..

[12] Lamberton, D., et B. Lapeyre, 1991, Introduction au Calcul Stochastique Appliqué à la Finance, Ellipses, Paris (英译本：Lamberton, D., & B. Lapeyre, 1996, Introduction to Stochastic Applied to Finance, Chapman & Hall, London).

[13] Lo, A. W. & A. C. MacKinlay, 1988, Stock market prices do not follow random walks: evidence from a simple specification test, *Review of Financial Study*, 1: 41–66.

[14] Merton, R., 1973, The theory of rational option pricing, *Bell Journal of Economics and Management Science,* **4:** 141–183.

[15] Merton, R. C., 1992, Continuous-Time Finance, Rev. ed., Blackwell, Oxford.

[16] Vasicek, O. A., 1977, An equilibrium characterization of the term structure, *J. Fin. Econ.* 5: 177–188..

附录：Black–Scholes 方程的求解

Black–Scholes 方程是指下列方程：

$$\frac{\partial u(t,x)}{\partial t}+\frac{\sigma^2x^2}{2}\frac{\partial^2u(t,x)}{\partial x^2}+rx\frac{\partial u(t,x)}{\partial x}-ru(t,x)=0. \tag{11}$$

欧式买入期权的定价问题就是对这个方程在下列终端条件下求解：

$$u(T,x)=(x-K)_+=\max\left[0,x-K\right]. \tag{12}$$

(11) 这样类型的方程称为抛物型方程。一般情况下，这种方程很难求得解析表达式。幸运的是，方程 (11) 有点特别，它对终端条件 (12) 的解刚好可以求出解析表达式，使得 Black–Scholes 公式得以问世。

最简单的抛物型方程 (它称为“热传导方程”、“热方程”或“扩散方程”) 有这样的形式：

$$\frac{\partial y(s,v)}{\partial s}-\frac{\partial^2 y(s,v)}{\partial v^2}=0, \qquad y(0,v)=y_0(v). \tag{13}$$

这个方程的解可以用积分表达：

$$\begin{aligned} y(s,v) &= \frac{1}{2\sqrt{\pi v}}\int_{-\infty}^{+\infty} y_0(v-w)\exp\left(-\frac{w^2}{4s}\right)dw \\ &= \frac{1}{2\sqrt{\pi v}}\int_{-\infty}^{+\infty} y_0(w)\exp\left(-\frac{(v-w)^2}{4s}\right)dw. \end{aligned} \tag{14}$$

对 (11), (12) 的求解要利用这一解的表达式。为此，设

$$u(t,x)=f(t)y(s(t),v(t,x)), \tag{15}$$

其中 $f(t)$, $s(t)$, $v(t,x)$ 和 $y(s,v)$ 都是待定的未知函数，并且假定 $y=y(s,v)$ 满足扩散方程 (13)。于是

$$\frac{\partial u}{\partial t}=f'(t)y(s,v)+f(t)\frac{\partial y}{\partial s}s'(t)+f(t)\frac{\partial y}{\partial v}\frac{\partial v}{\partial t}, \tag{16}$$

$$\frac{\partial u}{\partial x}=f(t)\frac{\partial y}{\partial v}\frac{\partial v}{\partial x}, \tag{17}$$

$$\frac{\partial^2 u}{\partial x^2}=f(t)\left[\frac{\partial^2 y}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2+\frac{\partial y}{\partial v}\frac{\partial^2 v}{\partial x^2}\right]. \tag{18}$$

把这些偏导数代入 Black–Scholes 方程 (11)，我们得到

$$\begin{aligned} 0 &= f'(t)y(s,v)+f(t)\left[\frac{\partial y}{\partial s}s'(t)+\frac{\partial y}{\partial v}\frac{\partial v}{\partial t}\right] \\ &\quad +\frac{\sigma^2x^2}{2}f(t)\left[\frac{\partial^2 y}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2+\frac{\partial y}{\partial v}\frac{\partial^2 v}{\partial x^2}\right] \\ &\quad +rxf(t)\frac{\partial y}{\partial v}\frac{\partial v}{\partial x}-rf(t)y(s,v). \end{aligned} \tag{19}$$

假设 $f=f(t)$ 满足 $f'(t)=rf(t)$, $f(T)=1$, 即 $f(t)=e^{-r(T-t)}$. 那么上述方程中的 f 都可提出来，使得 (19) 可简化为

$$\frac{\partial y}{\partial v}\frac{\partial v}{\partial t}+\frac{\partial y}{\partial s}s'(t)+\frac{\sigma^2x^2}{2}\left[\frac{\partial^2 y}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2+\frac{\partial y}{\partial v}\frac{\partial^2 v}{\partial x^2}\right]+rx\frac{\partial y}{\partial v}\frac{\partial v}{\partial x}=0. \tag{20}$$

又由 $y = y(s, v)$ 满足 (13), 我们又可假设

$$s'(t) = -a^2 = -\frac{\sigma^2 x^2}{2} \left(\frac{\partial v}{\partial x} \right)^2, \tag{21}$$

使得上述方程再少掉两项。由此再假设 $s(T) = 0$, 可得

$$s(t) = a^2(T - t). \tag{22}$$

这样一来，方程转化为 v 的如下方程：

$$\frac{\sigma^2 x^2}{2} \frac{\partial^2 v}{\partial x^2} + rx \frac{\partial v}{\partial x} + \frac{\partial v}{\partial t} = 0. \tag{23}$$

利用 a^2 的表达式 (21)，我们可求得

$$\frac{\partial v}{\partial x} = \sqrt{2} \left(\frac{a}{\sigma x} \right), \tag{24}$$

或

$$v(t, x) = \sqrt{2} \left(\frac{a}{\sigma} \right) \log \left(\frac{x}{K} \right) + b(t). \tag{25}$$

代入 (23) 得到

$$\frac{\sigma^2 x^2}{2} \left(-\frac{1}{x^2} \right) \sqrt{2} \left(\frac{a}{\sigma} \right) + rx\sqrt{2} \left(\frac{a}{\sigma} \right) \left(\frac{1}{x} \right) + b'(t) = 0, \tag{26}$$

或

$$b'(t) = \frac{1}{2} a\sigma\sqrt{2} - r\sqrt{2} \left(\frac{a}{\sigma} \right). \tag{27}$$

这样，假设 $b(T) = 0$, 就可得

$$b(t) = \sqrt{2} \left(\frac{a}{\sigma} \right) \left[r - \frac{\sigma^2}{2} \right] (T - t) \tag{28}$$

它使得

$$v(t, x) = \frac{a\sqrt{2}}{\sigma} \left[\log \left(\frac{x}{K} \right) + \left(r - \frac{\sigma^2}{2} \right) (T - t) \right] \tag{29}$$

现在利用 $u = u(t, x)$ 的表达式，可得

$$u(x, T) = f(T) y \left(s(T), v(x, T) \right) = y \left(0, \frac{a\sqrt{2}}{\sigma} \log \left(\frac{x}{K} \right) \right). \tag{30}$$

它必须满足

$$u(x,T)=y(0,k)=\begin{cases} K\left[\exp\left(\dfrac{k\sigma}{a\sqrt{2}}\right)-1\right], & \text{当 } k\geqslant 0,\\ 0, & \text{其他情形，}\end{cases} \tag{31}$$

其中

$$k=\frac{a\sqrt{2}}{\sigma}\log\left(\frac{x}{K}\right), \qquad x=K\exp\left(\frac{\sigma k}{a\sqrt{2}}\right). \tag{32}$$

应用这个条件，解的表达式 (14) 变为

$$y(s,v)=\frac{K}{2\sqrt{\pi s}}\int_{-\infty}^{v}\left[\exp\left(\frac{(v-w)\sigma}{a\sqrt{2}}\right)-1\right]\exp\left(-\frac{w^2}{4s}\right)dw. \tag{33}$$

令 $w=-q\sqrt{2s}$，它又变为

$$y(s,v)=\frac{K}{\sqrt{2\pi}}\int_{-v/\sqrt{2s}}^{\infty}\left[\exp\frac{\sigma\left(v+q\sqrt{2s}\right)}{a\sqrt{2}}-1\right]\exp\left(-\frac{q^2}{2}\right)dq. \tag{34}$$

利用方程 (22) 和 (29), 我们得到

$$\begin{aligned}\frac{v}{\sqrt{2s}} &= \frac{a\sqrt{2}}{\sigma}\left[\log\left(\frac{x}{K}\right)+\left(r-\frac{\sigma^2}{2}\right)(T-t)\right]\left[\frac{1}{a\sqrt{2}\sqrt{T-t}}\right]\\ &= \frac{\log\left(\dfrac{x}{K}\right)+\left(r-\dfrac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}=d_2. \end{aligned}\tag{35}$$

以及

$$\begin{aligned}&\left(\frac{\sigma}{a\sqrt{2}}\right)\left(v+q\sqrt{2s}\right)\\ &= \frac{\sigma}{a\sqrt{2}}\left\{\left(\frac{a\sqrt{2}}{\sigma}\right)\left[\log\left(\frac{x}{K}\right)+\left(r-\frac{\sigma^2}{2}\right)(T-t)\right]+qa\sqrt{2}\sqrt{T-t}\right\}\\ &= \log\left(\frac{x}{K}\right)+\left(r-\frac{\sigma^2}{2}\right)(T-t)+q\sigma\sqrt{t-T}.\end{aligned}\tag{36}$$

把这些值代入 (34)，我们有

$$\begin{aligned}&y(s(t),v(t,x))\\ &= \frac{K}{\sqrt{2\pi}}\int_{-d_2}^{+\infty}\exp\left[\log\left(\frac{x}{K}\right)+\left(r-\frac{\sigma^2}{2}\right)(T-t)+q\sigma\sqrt{T-t}-\frac{q^2}{2}\right]dq\\ &\quad -\frac{K}{\sqrt{2\pi}}\int_{-d_2}^{+\infty}\exp\left(-\frac{q^2}{2}\right)dq.\end{aligned}\tag{37}$$

它也可以记为

$$\begin{aligned} u(t,x) &= f(t)y(s(t),v(t,x)) \\ &= \frac{x\exp\left[\frac{-\sigma^2}{2}(T-t)\right]}{\sqrt{2\pi}}\int_{-d_2}^{+\infty}\exp(q\sigma\sqrt{T-t})\exp\left(-\frac{q^2}{2}\right)dq \\ &\quad -\frac{Ke^{-r(T-t)}}{\sqrt{2\pi}}\int_{-d_2}^{+\infty}\exp\left(-\frac{q^2}{2}\right)dq. \end{aligned} \tag{38}$$

再作变量替换 $p=-q$，那么有

$$u(t,x)=\frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d_2}\exp\left(\frac{-\sigma^2}{2}(T-t)-p\sigma\sqrt{T-t}-\frac{p^2}{2}\right)dp-Ke^{-r(T-t)}N(d_2). \tag{39}$$

令 $z=p+\sigma\sqrt{T-t}$，则

$$\begin{aligned} &\frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d_2}\exp\left(\frac{-\sigma^2}{2}(T-t)-p\sigma\sqrt{T-t}-\frac{p^2}{2}\right)dp \\ =&\frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d_2}\exp\left(-\frac{1}{2}\left(p+\sigma\sqrt{T-t}\right)^2\right)dp \\ =&\frac{x}{\sqrt{2\pi}}\int_{-\infty}^{d_2+\sigma\sqrt{T-t}}\exp\left(-\frac{z^2}{2}\right)dz=xN(d_1). \end{aligned} \tag{40}$$

因此，

$$u(t,x)=xN(d_1)-Ke^{-r(T-t)}N(d_2). \tag{41}$$

其中

$$d_1=\frac{\log\left(\frac{x}{K}\right)+\left(r+\frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \tag{42}$$

$$d_2=\frac{\log\left(\frac{x}{K}\right)+\left(r-\frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}, \tag{43}$$

$$N(d)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d}\exp\left(-\frac{1}{2}z^2\right)dz. \tag{44}$$

这样，最后就解得欧式买入期权的 Black–Scholes 公式。